

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 30, 2003, 14:34:49 ; Search time 16 Seconds
(without alignments)
1197.228 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161

Perfect score: 208
Sequence: 1 atggctgagctcttttgcctt.....gccctgagtgctgctact 111

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 656510

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09513999/runat_30042003_143404_25670/app-query.fasta.1.263
-DB=Published Applications_AA -QMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPT=0 -LOOPTXT=0 -UNITS=bits -SPART=1 -END=-1 -MATRIX=Dlosum62
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09513999_@CGN_1_1_17_@runat_30042003_143404_25670
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCOT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCOTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	29.8	4842	9	US-10-184-644-289
2	62	29.8	4842	9	US-10-184-634-289
3	60.5	29.1	77	9	US-09-965-528-1
4	60	28.8	695	9	US-10-184-644-567

5	60	28.8	695	9	US-10-184-634-567	Sequence 567, App
6	59	28.4	1768	9	US-10-184-644-343	Sequence 343, App
7	59	28.4	1768	9	US-10-184-634-343	Sequence 343, App
8	59	28.4	1941	9	US-10-123-155-165	Sequence 165, App
9	59	28.4	3162	9	US-10-123-155-111	Sequence 111, App
10	59	28.4	3192	9	US-10-123-155-75	Sequence 75, App
11	58.5	29.0	126	9	US-09-738-626-6097	Sequence 6097, App
12	58	27.9	2334	9	US-10-123-155-129	Sequence 129, App
13	58	27.9	2336	9	US-10-123-155-383	Sequence 383, App
14	57	28.2	950	10	US-09-823-355-9	Sequence 9, App
15	57	27.4	2476	9	US-10-184-644-585	Sequence 585, App
16	57	27.4	2476	9	US-10-184-634-585	Sequence 585, App
17	57	27.4	3951	9	US-10-184-644-119	Sequence 119, App
18	57	27.4	3951	9	US-10-184-634-119	Sequence 119, App
19	56.5	27.2	61	9	US-09-809-391-398	Sequence 398, App
20	56.5	27.2	1679	9	US-10-123-155-375	Sequence 375, App
21	56	27.7	56	9	US-09-796-692-748	Sequence 748, App
22	56	27.7	56	9	US-10-040-862-748	Sequence 748, App
23	56	26.9	80	10	US-09-864-761-43395	Sequence 43395, A
24	56	26.9	190	10	US-10-125-540-305	Sequence 305, App
25	56	26.9	190	10	US-09-764-870-305	Sequence 305, App
26	56	26.9	396	10	US-09-215-450-222	Sequence 22, App
27	56	26.9	396	10	US-09-953-956-13	Sequence 13, App
28	56	26.9	396	12	US-10-073-912-12	Sequence 12, App
29	56	26.9	1666	9	US-10-114-464-13	Sequence 13, App
30	56	26.9	2148	9	US-10-184-644-507	Sequence 507, App
31	56	26.9	2148	9	US-10-184-634-507	Sequence 507, App
32	56	26.9	3322	9	US-10-184-644-489	Sequence 489, App
33	56	26.9	3322	9	US-10-184-634-489	Sequence 489, App
34	56	26.9	1836	9	US-10-073-912-17	Sequence 17, App
35	55.5	26.7	147	9	US-10-042-296-6	Sequence 6, App
36	55.5	26.7	1572	9	US-10-184-644-65	Sequence 65, App
37	55.5	26.7	1572	9	US-10-184-634-65	Sequence 65, App
38	55.5	26.7	2037	9	US-10-184-634-13	Sequence 13, App
39	55.5	26.7	2037	9	US-10-184-634-13	Sequence 13, App
40	55	27.2	586	9	US-09-764-968-624	Sequence 624, App
41	55	27.2	586	9	US-09-955-956-64	Sequence 64, App
42	55	26.4	1738	9	US-10-184-644-239	Sequence 239, App
43	55	26.4	1738	9	US-10-184-634-239	Sequence 239, App
44	55	26.4	1968	9	US-10-123-155-163	Sequence 163, App
45	55	26.4	2806	9	US-10-123-155-201	Sequence 201, App

ALIGNMENTS

RESULT 1
US-10-184-644-289
; Sequence 289, Application US/10184644
; Publication No. US20030044930A1
GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION removed - See File Wrapper or Palm
; SEQ ID NO 289
; LENGTH: 4842
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-289


```
; LENGTH: 695
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-567

Alignment Scores:
Pred. No.: 20.9      Length: 695
Score: 60.00        Matches: 13
Percent Similarity: 48.39%      Conservative: 2
Best Local Similarity: 41.94%    Mismatches: 15
Query Match: 28.85%             Indels: 1
DB: 9                       Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-184-644-567 (1-695)

OY 15 TGCCCTGAGAGATCTTTTCATCTTGCAGGACTTCGGCGCCGAGATGTAAACT 74
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 606 CysCysAlaGlyGlyAlaCysCysAlaAlaGlyAlaAlaGlyThrGlyAlaCysGly 625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 75 CCTGGCTCTGTGTGTGCTGAGTGGCTGC 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 ThrGlyGly-CysThrCysCysAlaGlyCys 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
; Sequence 567, Application US/10184634
; Publication No. US2003006864A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 567
; LENGTH: 695
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-567

Alignment Scores:
Pred. No.: 20.9      Length: 695
Score: 60.00        Matches: 13
Percent Similarity: 48.39%      Conservative: 2
Best Local Similarity: 41.94%    Mismatches: 15
Query Match: 28.85%             Indels: 1
DB: 9                       Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-184-634-567 (1-695)

OY 15 TGCCCTGAGAGATCTTTTCATCTTGCAGGACTTCGGCGCCGAGATGTAAACT 74
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 606 CysCysAlaGlyGlyAlaCysCysAlaAlaGlyAlaAlaGlyThrGlyAlaCysGly 625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 75 CCTGGCTCTGTGTGTGCTGAGTGGCTGC 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 ThrGlyGly-CysThrCysCysAlaGlyCys 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
; Sequence 343, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 343
; LENGTH: 1768
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-343

Alignment Scores:
Pred. No.: 30.9      Length: 1768
Score: 59.00        Matches: 13
Percent Similarity: 50.00%      Conservative: 3
Best Local Similarity: 40.62%    Mismatches: 9
Query Match: 28.37%             Indels: 7
DB: 9                       Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-184-644-343 (1-1768)

OY 15 TGCCCTGAGAGATCTTTTCATCTTGCAGGACTTCGGCGCCGAGATGTAAACT 74
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 CysCysAlaGly-----GlyAlaCysGlyGlyGlyCysGlyThr 572
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 75 CCTGGCTCTGTGTGTGCTGAGTGGCTGC 108
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 GlyGlyAla-CysAlaCysCysThrGlyCysThr 583
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
; Sequence 343, Application US/10184634
; Publication No. US2003006864A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 343
; LENGTH: 1768
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-343
```

```
Alignment Scores:
Pred. No.: 30.9 Length: 1768
Score: 59.00 Matches: 13
Percent Similarity: 50.00% Conservative: 3
Best Local Similarity: 40.62% Mismatches: 9
Query Match: 28.37% Indels: 7
DB: 9 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-184-634-343 (1-1768)
QY 15 TGCGCTGCAGAGATCTTTTCATCTTTCAGGAGCTTCGCGCCGAGTATGAAACT 74
DB 559 CysCysAlaGly-----GlyAlaCysGlyGlyGlyCysCysGlyThr 572
QY 75 CCGGCTCTGTGTGTGCGCTGAGTGGCTGCTCT 108
DB 573 GlyGlyAla-CysAlaCysCysThrGlyCysThr 583

RESULT 8
US-10-123-155-165
Sequence 165, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C30
CURRENT APPLICATION NUMBER: US/10/123,155
PRIORITY FILING DATE: 2002-04-15
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 165
LENGTH: 1941
TYPE: DNA
ORGANISM: Homo Sapien
US-10-123-155-165

Alignment Scores:
Pred. No.: 31.2 Length: 1941
Score: 59.00 Matches: 11
Percent Similarity: 60.87% Conservative: 3
Best Local Similarity: 47.83% Mismatches: 8
Query Match: 28.37% Indels: 1
DB: 9 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-123-155-165 (1-1941)
QY 45 GGGAGCTTCGGGCGCGGAGTATGFAAAACCTGCTCTGTGTGCTGCTGAGTGGCTG 104
DB 1815 GlyCysThrGlyGlyGlyCysCysThrAlaGlyGly-CysThrCysThrThrGlyCys 1834
QY 105 CTCTACT 111
DB 1834 sThrThr 1836

RESULT 9
US-10-123-155-111
```

```
; Sequence 111, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

US-10-123-155-75
Sequence 75, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-123-155-111 (1-3162)
QY 3 GGGTGAATCTTTTGGCTTCAGAGATCTTTTCATCTTTCAGGAGCTTCGGGCGGGA 62
DB 1320 GlyThrCysCysCysAlaGlyAlaGlyCysCysThrGlyGlyGlyAlaGly 1339
QY 63 GATGTAAACCTCGGCTCTGTGTGCTGCGCTGAGTGGCTGCTCTACT 111
DB 1340 CysCysGlyThrAlaGlyGlyCys-CysCysAlaAlaCysCysThr 1355

RESULT 10
US-10-123-155-75
Sequence 75, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin

Alignment Scores:
Pred. No.: 32.8 Length: 3162
Score: 59.00 Matches: 15
Percent Similarity: 43.24% Conservative: 1
Best Local Similarity: 40.54% Mismatches: 20
Query Match: 28.37% Indels: 1
DB: 9 Gaps: 0
```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123.155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 75
; LENGTH: 3192
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1428,1431
; OTHER INFORMATION: unknown base
; US-10-123-155-75

Alignment Scores:
Pred. No.: 32.9 Length: 3192
Score: 59.00 Matches: 13
Percent Similarity: 53.12% Conservative: 4
Best Local Similarity: 40.62% Mismatches: 14
Query Match: 28.37% Indels: 1
DB: Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-123-155-75 (1-3192)
QY 15 TGGCTTCAGAGATCTTTTCATCTTTCAGAGGACCTCGGCGGAGTATGTAACACT 74
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1554 CyscysalaglYalaglYalaaLaacysalaglYalaglYalaglYalaglYalaglY 1573
QY 75 CCGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 108
    ::||| ||| ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1574 ThrThrAla-CysAlaCysCysThrGlYcysAla 1584

RESULT 11
US-09-738-626-6097
; Sequence 6097, Application US/09/738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6097
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6097

Alignment Scores:
Pred. No.: 27.1 Length: 126
Score: 58.50 Matches: 12
Percent Similarity: 59.26% Conservative: 4
Best Local Similarity: 44.44% Mismatches: 8
```

```

Query Match: 28.96% Indels: 3
DB: 9 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-738-626-6097 (1-126)
QY 91 ACACACAGAGACCCGAGGTTTTCATATCTCGGCGCCAGAGTCCCT-----GCA 41
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 90 SerHisHisLysAspGlyLeuThrArgLeuProAlaProIleuProAsnAlaIle 109
QY 40 AAGATGAAAAAGAAATCTGCA 20
    ||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110 LysAlaArgLysAsnProAla 116

RESULT 12
US-10-123-155-129
; Sequence 129, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tunas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123.155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 129
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-10-123-155-129

Alignment Scores:
Pred. No.: 42.6 Length: 2334
Score: 58.00 Matches: 13
Percent Similarity: 58.33% Conservative: 1
Best Local Similarity: 54.17% Mismatches: 7
Query Match: 27.88% Indels: 3
DB: Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-123-155-129 (1-2334)
QY 42 GCAGGAGACTTCGGGGCGGAGATATGTAACACTCCTGGGCTCTGTGTGCTGCTGATGG 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 277 ATAGTThnGlyGlyAlaGlyCysCys-----ThrGlyAla-CysAlaCysThrGlyG 294
QY 102 CGCTCTACT 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 YCysGlyThr 297

RESULT 13
US-10-123-155-383
; Sequence 383, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
```

```

; SEQ ID NO 9
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20010025098A1 1794154
US-09-823-356-9

Alignment Scores:
Pred. NO.:          52           Length:      950
Score:              57.00        Matches:     15
Percent Similarity: 58.62%       Conservative: 2
Best Local Similarity: 51.72%    Mismatches:  8
Query Match:        28.22%       Indels:      4
DB:                 10           Gaps:        2

US-09-513-999C-3792_COPY_51_161 (1-111) x US-09-823-356-9 (1-950)
QY      88 CACGAGAGCCAGGAGGTTCATACATCCGGCCCCAGAAAGTCCCTCGCAAGAAGTGAAA--- 32
Db      475 HisagagapPProanValVal-----ProllyProProlysProalatlatsGIuLysPro 492
QY      31 --AAGATCCTCGCAGAGCAAAAGATC 8
Db      493 ProlysLysLysAlaGlnAspLysIle 501

RESULT 15
US-10-184-644-585
; Sequence 585, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430RIC227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 585
; LENGTH: 2476
; TYPE: DNA
; ORGANISM: Homo Saplen
US-10-184-644-585

Alignment Scores:
Pred. NO.:          57.5         Length:     2476
Score:             57.00         Matches:     13
Percent Similarity: 50.00%       Conservative: 2
Best Local Similarity: 43.33%    Mismatches:  8
Query Match:       27.40%       Indels:      7
DB:                9           Gaps:        1

US-09-513-999C-3792_COPY_51_161 (1-111) x US-10-184-644-585 (1-2476)
QY      42 GCAGGAGACTCTGCGGGGCCGAGATGATAAACCTCCGGG-----TCT 83
Db      60 AlGelYThrlYlglYglYalaCysAlatHrThrgLYcSThrElYalacYSalaLa 79
QY      84 CTGTGTGTGCCTAGTGGTGTCTACT 111
Db      80 ThfCYs-CVscvsglYgIvcvstHrThr 88

```

Wed Apr 30 15:10:00 2003

us-09-513-999C-3792_copy_51_161.n2p.rapb

Page 7

Search completed: April 30, 2003, 14:37:30
Job time : 21 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 30, 2003, 14:35:15 ; Search time 12 Seconds
(without alignments)
544.324 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161

Perfect score: 208
Sequence: 1 atgagtgatcttttgcctt.....gcctcagtgctgctctact 111

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+.n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09513999/rnat_30042003_143402_25630/app_query.fasta_1.263
-DB=Issued Patents_AA -QFMT=fastn -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09513999_ECGN_1_1_28_@rnat_30042003_143402_25630 -NCPD=6 -ICPD=3
-NO_XLPEXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-MARN.TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUTS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	63.5	31.4	640	4	US-09-262-773-4
C 2	63.5	31.4	648	4	US-09-262-773-2
C 3	60.5	29.1	218	2	US-08-399-889-25
C 4	60.5	29.1	218	3	US-09-167-364-25
C 5	60.5	29.1	218	4	US-09-439-897-4
C 6	60.5	29.1	268	4	US-09-589-927-6
C 7	60.5	29.1	268	4	US-09-277-665-6
C 8	60.5	29.1	471	2	US-08-399-889-24
C 9	60.5	29.1	471	3	US-09-167-364-24
C 10	60.5	29.1	471	4	US-09-439-897-2
C 11	59	29.2	1297	2	US-08-290-731c-4
C 12	56.5	27.2	61	4	US-09-149-476-398

13	56	26.9	349	4	US-09-032-523-3	Sequence 3, Appl1
14	56	26.9	336	1	US-08-208-007A-13	Sequence 13, Appl1
15	56	26.9	326	4	US-09-032-523-9	Sequence 9, Appl1
16	56	26.9	336	4	US-08-515-095A-13	Sequence 13, Appl1
17	56	26.9	336	4	US-08-798-096-13	Sequence 13, Appl1
18	56	26.9	336	4	US-08-798-095A-13	Sequence 13, Appl1
19	54	26.0	1720	2	US-08-477-451-12	Sequence 12, Appl1
C 20	52.5	26.0	976	3	US-08-560-005-2	Sequence 2, Appl1
C 21	52.5	26.0	976	3	US-09-195-868-14	Sequence 14, Appl1
C 22	52.5	26.0	976	4	US-09-418-540-2	Sequence 2, Appl1
C 23	52.5	26.0	1187	4	US-08-664-962B-8	Sequence 8, Appl1
C 24	52.5	26.0	1187	3	US-09-311-743-8	Sequence 8, Appl1
C 25	52.5	26.0	1189	4	US-09-195-868-15	Sequence 15, Appl1
C 26	52.5	26.0	1229	3	US-09-195-868-28	Sequence 28, Appl1
C 27	52	25.0	418	2	US-08-494-807-18	Sequence 18, Appl1
C 28	52	25.0	418	5	PCT-US96-10986-18	Sequence 18, Appl1
C 29	52	25.0	2088	4	US-09-548-372D-13	Sequence 13, Appl1
C 30	52	25.0	2088	4	US-09-548-367D-13	Sequence 13, Appl1
C 31	51.5	24.8	396	2	US-09-061-337-12	Sequence 12, Appl1
C 32	51.5	24.8	396	2	US-09-122-129-12	Sequence 12, Appl1
C 33	51.5	24.8	396	3	US-09-340-991-12	Sequence 12, Appl1
C 34	51.5	24.8	396	4	US-08-974-609-12	Sequence 12, Appl1
C 35	51.5	24.8	396	4	US-09-549-098-12	Sequence 12, Appl1
C 36	51.5	24.8	1417	4	US-08-900-230-3	Sequence 3, Appl1
C 37	51	24.5	24	4	US-09-443-501A-21	Sequence 21, Appl1
C 38	51	25.2	1119	4	US-09-396-651B-2	Sequence 2, Appl1
C 39	50.5	25.0	564	1	US-07-872-644-53	Sequence 53, Appl1
C 40	50.5	25.0	564	1	US-08-297-494-53	Sequence 53, Appl1
C 41	50.5	25.0	564	1	US-08-297-510-53	Sequence 53, Appl1
C 42	50.5	25.0	564	1	US-08-479-532-53	Sequence 53, Appl1
C 43	50.5	25.0	564	1	US-08-455-526-53	Sequence 53, Appl1
C 44	50.5	25.0	564	1	US-08-455-525-53	Sequence 53, Appl1
C 45	50.5	25.0	564	3	US-09-139-491-53	Sequence 53, Appl1

ALIGNMENTS

RESULT 1
US-09-262-773-4
; Sequence 4, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: MYriad 3
; CURRENT APPLICATION NUMBER: US/09/262, 773
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 640
; TYPE: PRT
; ORGANISM: human
US-09-262-773-4
Alignment Scores:
Pred. No.: 1.15
Score: 63.50
Percent Similarity: 66.67%
Best Local Similarity: 44.44%
Query Match: 31.44%
DB: 4
US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-262-773-4 (1-640)
QY 110 GAGAGCAGCCACCTGAGGCACACAGAGACCCAG-----AGTTTACTACTACCGCG 57
DB 292 IlegInguPro--GInGuInThrInGuProGInGuInLeuSerPheThyTThrely 310

```
OY 56 CCCAGACTCCCTGCAAGATGAAAGAAATCTGCAAGGCAAAAGAT 9
||||| |||||||::: ||| ::|
Db 311 AspArgSer-----LysAspGluGluGluGluGluGluGluGluGlu 324

RESULT 2
US-09-262-773-2
; Sequence 2, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 648
; TYPE: PRT
; ORGANISM: human
US-09-262-773-2

Alignment Scores:
Pred. No.: 1.15 Length: 648
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5
DB: 4 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-262-773-2 (1-648)
OY 110 GTACGAGCGACGACACGACGACAGACCCAGG-----AGTTTACATCTCCGCG 57
|||||::: ||||| ||| |||||||::: |||
Db 300 lIeInGluPro--GInGluThGInGluProGluIleuSerPheTrItyrInGly 318

OY 56 CCCAGAGTCCCTGCAAGATGAAAGAAATCTGCAAGGCAAAAGAT 9
||||| |||||||::: ||| ::|
Db 319 AspArgSer-----LysAspGluGluGluGluGluGluGluGluGlu 332

RESULT 3
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 25
LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-399-889-25

Alignment Scores:
Pred. No.: 2.49 Length: 218
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 2 Gaps: 3
```

```
US-09-513-999c-3792_COPY_51_161 (1-111) x US-08-399-889-25 (1-218)
OY 3 GGGTGATCTTTTGGCTTCAGAGATCTTTTCATCTTT-----GCAGGACTTCT 53
||||| ||||||| ||||||| |||||||
Db 121 GYTRPILserLeuTrpYsglyPheSerPheIleMetPheThrSerAlaIglySerIu 140

OY 54 GGGGCCGGA---GTATGTAAGTCTCGGCTCTGTGTGTGCTGAGTGG 101
||||| |||||||::: |||::: |||:::
Db 141 GYAlaIglyGlnAlaIleuAlaSerProGlySer---CysLeuGluGluPhe 156

RESULT 4
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 25
LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Alignment Scores:
Pred. No.: 2.49 Length: 218
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 3 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-167-364-25 (1-218)
OY 3 GGGTGATCTTTTGGCTTCAGAGATCTTTTCATCTTT-----GCAGGACTTCT 53
||||| ||||||| ||||||| |||||||
Db 121 GYTRPILserLeuTrpYsglyPheSerPheIleMetPheThrSerAlaIglySerIu 140

OY 54 GGGGCCGGA---GTATGTAAGTCTCGGCTCTGTGTGTGCTGAGTGG 101
||||| |||||||::: |||::: |||:::
Db 141 GYAlaIglyGlnAlaIleuAlaSerProGlySer---CysLeuGluGluPhe 156

RESULT 5
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 627558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Alignment Scores:
Pred. No.: 2.49 Length: 218
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
```



```

?      TYPE: PRT
?      ORGANISM: Calf
US-09-167-364-24

Alignment Scores:
Pred. No.:          2.81
Score:              60.50
Percent Similarity: 56.76%
Best Local Similarity: 45.95%
Query Match:       29.09%
DB:                3

US-09-513-999C-3792.COPY_51_161 (1-111) x US-09-167-364-24 (1-471)
OY      3 GGGGGAATCTTTGGCCTTGACGATCTTTTTCATCTTT-----GCAGGACTTCT 53
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      374 GATPTLlesterleutPTlysglyPheSerPheIleMetPheTherIaAglyserglu 39
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY      54 GGGGCCGGA---GTATGTAACATCTCGGCTCTCTGTGTCGCTGAGTGG 101
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
          ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
          394 GATyAlaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGlnIuphe 409
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 10
US-09-439-897-2
Sequence 2, Application US/09439897
Patent No. 6277558
GENERAL INFORMATION:
APPLICANT: HUDSON, BILLY G
FILE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 95-1263-C
CURRENT APPLICATION NUMBER: US/09/439,897
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 471
TYPE: PRT
ORGANISM: Bos taurus
US-09-439-897-2

Alignment Scores:
Pred. No.:          2.81
Score:              60.50
Percent Similarity: 56.76%
Best Local Similarity: 45.95%
Query Match:       29.09%
DB:                4

S-09-513-999C-3792.COPY_51_161 (1-111) x US-09-439-897-2 (1-471)
OY      3 GGGTGAATCTTTGGCCTTGACGATCTTTTTCATCTTT-----GCAGGACTTCT 53
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      374 GATPTLlesterleutPTlysglyPheSerPheIleMetPheTherIaAglyserglu 39
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY      54 GGGGCCGGA---GTATGTAACATCTCGGCTCTCTGTGTCGCTGAGTGG 101
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
          ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
          394 GATyAlaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGlnIuphe 409
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 11
US-08-290-731C-4
Sequence 4, Application US/08290731C
Patent No. 5843646
GENERAL INFORMATION:
APPLICANT: BOWTELL, David Douglas Lawrence
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
TITLE OF INVENTION: SON OF SEVENTLESS (MSOS) GENE,
TITLE OF INVENTION: AND MSOS POLYPEPTIDES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MION, ZINN, MACPHEAR & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.

```

```

1 COUNTRY: USA
2 ZIP: 20037
3 COMPUTER READABLE FORM:
4 MEDIUM TYPE: Floppy disk
5 COMPUTER: IBM PC compatible
6 OPERATING SYSTEM: PC-DOS/MS-DOS
7 SOFTWARE: Patentin Release #1.0, Version #1.25
8 CURRENT APPLICATION DATA:
9 APPLICATION NUMBER: US/08/290,731C
10 FILING DATE: 17-OCT-1994
11 CLASSIFICATION: 435
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: PCT/AU93/00068
14 FILING DATE: 17-FEB-1993
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: PI0921/92
17 FILING DATE: 17-FEB-1992
18 ATTORNEY/AGENT INFORMATION:
19 NAME: KIT, Gordon
20 REGISTRATION NUMBER: 30,764
21 REFERENCE/DOCKET NUMBER: Q-36066
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: (202) 293-7060
24 TELEFAX: (202) 293-7860
25 TELEX: 6491103
26 INFORMATION FOR SEQ ID NO: 4:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 1297 amino acids
29 TYPE: amino acid
30 TOPOLOGY: linear
31 MOLECULE TYPE: protein
32 US-08-290-731C-4
33
34 Alignment Scores:
35 Pred. No.: 5.27 Length: 1297
36 Score: 59.00 Matches: 13
37 Percent Similarity: 48.57% Conservative: 4
38 Best Local Similarity: 37.14% Mismatches: 12
39 Query Match: 29.21% Indels: 6
40 DB: 2 Gaps: 1
41
42 US-09-513-999C-3792_COPY_51_161 (1-111) x US-08-290-731C-4 (1-1297)
43
44 QY 106 AGCAGCCACGACGACGACACACAGACCAGAGTTTACATACCTCCGGGCCCCAAGATC 47
45 |||||
46 Db 1244 SerSerHisser-----SerLeuAlaHisLeuProAlaProProAl 1257
47
48 QY 46 CCTGCAAGAGTGA AAAAGAAATCTCTGCAAGCAAAAGATCCACCA 2
49 |||:::
50 Db 1258 ProProAlaGlnAsnSerSerProLeuLeuProLysLeuProPro 1272
51
52 RESULT 12
53 US-09-149-476-398
54 : Sequence 398, Application US/09149476
55 : Patent No. 6420526
56 : GENERAL INFORMATION:
57 : APPLICANT: Rosen et al.
58 : TITLE OF INVENTION: 186 Human Secreted proteins
59 : FILE REFERENCE: P2002P1
60 : CURRENT APPLICATION NUMBER: US/09/149,476
61 : EARLIER FILING DATE: 1998-09-08
62 : EARLIER APPLICATION NUMBER: PCT/US98/04493
63 : EARLIER FILING DATE: 1998-03-06
64 : EARLIER APPLICATION NUMBER: 60/040,162
65 : EARLIER FILING DATE: 1997-03-07
66 : EARLIER APPLICATION NUMBER: 60/040,333
67 : EARLIER FILING DATE: 1997-03-07
68 : EARLIER APPLICATION NUMBER: 60/038,621
69 : EARLIER FILING DATE: 1997-03-07
70 : EARLIER APPLICATION NUMBER: 60/040,626
71 : EARLIER FILING DATE: 1997-03-07
72 : EARLIER APPLICATION NUMBER: 60/040,334
73 : EARLIER FILING DATE: 1997-03-07

```

1	EARLIER APPLICATION NUMBER: 60/040,330
2	EARLIER FILING DATE: 1997-03-07
3	EARLIER APPLICATION NUMBER: 60/040,163
4	EARLIER FILING DATE: 1997-03-07
5	EARLIER APPLICATION NUMBER: 60/047,600
6	EARLIER FILING DATE: 1997-05-23
7	EARLIER APPLICATION NUMBER: 60/047,615
8	EARLIER FILING DATE: 1997-05-23
9	EARLIER APPLICATION NUMBER: 60/047,597
10	EARLIER FILING DATE: 1997-05-23
11	EARLIER APPLICATION NUMBER: 60/047,502
12	EARLIER FILING DATE: 1997-05-23
13	EARLIER APPLICATION NUMBER: 60/047,633
14	EARLIER FILING DATE: 1997-05-23
15	EARLIER APPLICATION NUMBER: 60/047,583
16	EARLIER FILING DATE: 1997-05-23
17	EARLIER APPLICATION NUMBER: 60/047,503
18	EARLIER FILING DATE: 1997-05-23
19	EARLIER APPLICATION NUMBER: 60/047,592
20	EARLIER FILING DATE: 1997-05-23
21	EARLIER APPLICATION NUMBER: 60/047,581
22	EARLIER FILING DATE: 1997-05-23
23	EARLIER APPLICATION NUMBER: 60/047,584
24	EARLIER FILING DATE: 1997-05-23
25	EARLIER APPLICATION NUMBER: 60/047,500
26	EARLIER FILING DATE: 1997-05-23
27	EARLIER APPLICATION NUMBER: 60/047,587
28	EARLIER FILING DATE: 1997-05-23
29	EARLIER APPLICATION NUMBER: 60/047,492
30	EARLIER FILING DATE: 1997-05-23
31	EARLIER APPLICATION NUMBER: 60/047,598
32	EARLIER FILING DATE: 1997-05-23
33	EARLIER APPLICATION NUMBER: 60/047,613
34	EARLIER FILING DATE: 1997-05-23
35	EARLIER APPLICATION NUMBER: 60/047,582
36	EARLIER FILING DATE: 1997-05-23
37	EARLIER APPLICATION NUMBER: 60/047,596
38	EARLIER FILING DATE: 1997-05-23
39	EARLIER APPLICATION NUMBER: 60/047,612
40	EARLIER FILING DATE: 1997-05-23
41	EARLIER APPLICATION NUMBER: 60/047,632
42	EARLIER FILING DATE: 1997-05-23
43	EARLIER APPLICATION NUMBER: 60/047,601
44	EARLIER FILING DATE: 1997-05-23
45	EARLIER APPLICATION NUMBER: 60/043,580
46	EARLIER FILING DATE: 1997-04-11
47	EARLIER APPLICATION NUMBER: 60/043,568
48	EARLIER FILING DATE: 1997-04-11
49	EARLIER APPLICATION NUMBER: 60/043,314
50	EARLIER FILING DATE: 1997-04-11
51	EARLIER APPLICATION NUMBER: 60/043,569
52	EARLIER FILING DATE: 1997-04-11
53	EARLIER APPLICATION NUMBER: 60/043,311
54	EARLIER FILING DATE: 1997-04-11
55	EARLIER APPLICATION NUMBER: 60/043,671
56	EARLIER FILING DATE: 1997-04-11
57	EARLIER APPLICATION NUMBER: 60/043,674
58	EARLIER FILING DATE: 1997-04-11
59	EARLIER APPLICATION NUMBER: 60/043,669
60	EARLIER FILING DATE: 1997-04-11
61	EARLIER APPLICATION NUMBER: 60/043,312
62	EARLIER FILING DATE: 1997-04-11
63	EARLIER APPLICATION NUMBER: 60/043,313
64	EARLIER FILING DATE: 1997-04-11
65	EARLIER APPLICATION NUMBER: 60/043,672
66	EARLIER FILING DATE: 1997-04-11
67	EARLIER APPLICATION NUMBER: 60/043,315
68	EARLIER FILING DATE: 1997-04-11
69	EARLIER APPLICATION NUMBER: 60/048,974

1	EARLIER	FILING DATE:	1997-06-06
2	EARLIER	APPLICATION NUMBER:	60/056, 886
3	EARLIER	FILING DATE:	1997-08-22
4	EARLIER	APPLICATION NUMBER:	60/056, 877
5	EARLIER	FILING DATE:	1997-08-22
6	EARLIER	APPLICATION NUMBER:	60/056, 889
7	EARLIER	FILING DATE:	1997-08-22
8	EARLIER	APPLICATION NUMBER:	60/056, 893
9	EARLIER	FILING DATE:	1997-08-22
10	EARLIER	APPLICATION NUMBER:	60/056, 630
11	EARLIER	FILING DATE:	1997-08-22
12	EARLIER	APPLICATION NUMBER:	60/056, 878
13	EARLIER	FILING DATE:	1997-08-22
14	EARLIER	APPLICATION NUMBER:	60/056, 662
15	EARLIER	FILING DATE:	1997-08-22
16	EARLIER	APPLICATION NUMBER:	60/056, 872
17	EARLIER	FILING DATE:	1997-08-22
18	EARLIER	APPLICATION NUMBER:	60/056, 882
19	EARLIER	FILING DATE:	1997-08-22
20	EARLIER	APPLICATION NUMBER:	60/056, 637
21	EARLIER	FILING DATE:	1997-08-22
22	EARLIER	APPLICATION NUMBER:	60/056, 903
23	EARLIER	FILING DATE:	1997-08-22
24	EARLIER	APPLICATION NUMBER:	60/056, 888
25	EARLIER	FILING DATE:	1997-08-22
26	EARLIER	APPLICATION NUMBER:	60/056, 879
27	EARLIER	FILING DATE:	1997-08-22
28	EARLIER	APPLICATION NUMBER:	60/056, 880
29	EARLIER	FILING DATE:	1997-08-22
30	EARLIER	APPLICATION NUMBER:	60/056, 894
31	EARLIER	FILING DATE:	1997-08-22
32	EARLIER	APPLICATION NUMBER:	60/056, 911
33	EARLIER	FILING DATE:	1997-08-22
34	EARLIER	APPLICATION NUMBER:	60/056, 636
35	EARLIER	FILING DATE:	1997-08-22
36	EARLIER	APPLICATION NUMBER:	60/056, 874
37	EARLIER	FILING DATE:	1997-08-22
38	EARLIER	APPLICATION NUMBER:	60/056, 910
39	EARLIER	FILING DATE:	1997-08-22
40	EARLIER	APPLICATION NUMBER:	60/056, 864
41	EARLIER	FILING DATE:	1997-08-22
42	EARLIER	APPLICATION NUMBER:	60/056, 631
43	EARLIER	FILING DATE:	1997-08-22
44	EARLIER	APPLICATION NUMBER:	60/056, 845
45	EARLIER	FILING DATE:	1997-08-22
46	EARLIER	APPLICATION NUMBER:	60/056, 892
47	EARLIER	FILING DATE:	1997-08-22
48	EARLIER	APPLICATION NUMBER:	60/057, 761
49	EARLIER	FILING DATE:	1997-08-22
50	EARLIER	APPLICATION NUMBER:	60/047, 595
51	EARLIER	FILING DATE:	1997-05-23
52	EARLIER	APPLICATION NUMBER:	60/047, 585
53	EARLIER	FILING DATE:	1997-05-23
54	EARLIER	APPLICATION NUMBER:	60/047, 586
55	EARLIER	FILING DATE:	1997-05-23
56	EARLIER	APPLICATION NUMBER:	60/047, 590
57	EARLIER	FILING DATE:	1997-05-23
58	EARLIER	APPLICATION NUMBER:	60/047, 534
59	EARLIER	FILING DATE:	1997-05-23
60	EARLIER	APPLICATION NUMBER:	60/047, 589
61	EARLIER	FILING DATE:	1997-05-23
62	EARLIER	APPLICATION NUMBER:	60/047, 593
63	EARLIER	FILING DATE:	1997-05-23
64	EARLIER	APPLICATION NUMBER:	60/047, 614
65	EARLIER	FILING DATE:	1997-05-23
66	EARLIER	APPLICATION NUMBER:	60/043, 578
67	EARLIER	FILING DATE:	1997-04-11
68	EARLIER	APPLICATION NUMBER:	60/043, 576
69	EARLIER	FILING DATE:	1997-04-11

PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
FILING DATE: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,7134
REFERENCE/DOCKET NUMBER: 325800-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-13

Alignment Scores:
Pred. No.: 11.2 Length: 396
Score: 56.00 Matches: 12
Percent Similarity: 62.07% Conservative: 6
Best Local Similarity: 41.38% Mismatches: 11
Query Match: 26.92% Indels: 0
Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-08-208-007A-13 (1-396)

QY 4 GGTGATCTTTTGCCTTCGAGGATCTTTTCATCTTTGAGGACTTCGGCCGAG 63
||| |||||:||||| ||||| |||||:||||| :|||
Db 132 GYGLNserPheSerIleGlnTyrGlyThrGlySerLeuSerGlyIleIleGlyAlaAsp 151

QY 64 TATGTAAACCTCCTGGCTCTCTGTG 90
||| :||| ||||| |||
Db 152 GlnValSerValGluGlyLeuThrVal 160

RESULT 15
US-09-032-523-9
Sequence 9, Application US/09032523
Patent No. 6232454
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl
APPLICANT: Baugh, Mariah
TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 181994
US-09-032-523-9

Alignment Scores:
Pred. No.: 11.2 Length: 396
Score: 56.00 Matches: 12
Percent Similarity: 62.07% Conservative: 6
Best Local Similarity: 41.38% Mismatches: 11
Query Match: 26.92% Indels: 0
Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-032-523-9 (1-396)

QY 4 GGTGATCTTTTGCCTTCGAGGATCTTTTCATCTTTGAGGACTTCGGCCGAG 63
||| |||||:||||| ||||| |||||:||||| :|||
Db 132 GYGLNserPheSerIleGlnTyrGlyThrGlySerLeuSerGlyIleIleGlyAlaAsp 151

QY 64 TATGTAAACCTCCTGGCTCTCTGTG 90
||| :||| ||||| |||
Db 152 GlnValSerValGluGlyLeuThrVal 160

Search completed: April 30, 2003, 14:39:10
Job time : 14 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 30, 2003, 14:34:46 ; Search time 27.5 Seconds
(without alignments)
1663.362 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161
Perfect score: 208
Sequence: 1 atggatgcatcttctgcctt.....gctcgtgagtcgtctact 111

Scoring table:
BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09513999/runatc_30042003_143401_25604/app_query.fasta_1.263
-DB=SPTRMBL_21 -OFMT=fastan -SUFFIX=n2p -ISPT -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09513999 @CGN_1.1.138 @runatc_30042003_143401_25604 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGESOURT -NRG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.potent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
c 1	66	32.7	83	15	Q80822	Q80822 human t-cel

c 2	63.5	31.4	547	4	Q9Y5A5	Q9Y5A5 homo sapien
c 3	63.5	31.4	634	4	Q9NSM4	Q9NSM4 homo sapien
c 4	60.5	29.1	203	6	Q28682	Q28682 oryctolagus
c 5	60.5	29.1	210	6	Q28273	Q28273 canis famill
c 6	60.5	29.1	644	5	Q9N4W3	Q9N4W3 caenorhabdi
c 7	60	28.8	314	5	Q20146	Q20146 caenorhabdi
c 8	60	29.7	533	11	P97320	P97320 mus musculu
c 9	60	28.8	1145	10	Q9LF10	Q9LF10 arabisdopsi
c 10	60	29.7	5170	5	Q17490	Q17490 caenorhabdi
c 11	60	29.7	6994	5	Q17343	Q17343 caenorhabdi
c 12	59.5	29.5	1167	5	Q9Y067	Q9Y067 theileria a
c 13	59	28.4	382	16	Q8UEA8	Q8UEA8 agrobacteri
c 14	59	29.2	537	11	Q9QY56	Q9QY56 mus musculu
c 15	58.5	28.1	230	11	Q63122	Q63122 rattus norv
c 16	58.5	28.1	231	15	Q56309	Q56309 wallie epi
c 17	58.5	29.0	283	4	Q8MM87	Q8MM87 homo sapien
c 18	58.5	29.0	583	5	Q9VYH5	Q9VYH5 drosophila
c 19	58	28.7	197	10	Q9SZ01	Q9SZ01 arabisdopsi
c 20	58	27.9	257	12	Q9J2L8	Q9J2L8 macaca mula
c 21	58	27.9	299	12	Q9MR74	Q9MR74 macaca mula
c 22	58	27.9	499	17	Q8RPF6	Q8RPF6 methanosarc
c 23	58	28.7	1099	5	Q9YMS5	Q9YMS5 drosophila
c 24	57.5	27.6	212	6	Q28567	Q28567 ovis aries
c 25	57.5	28.5	364	4	Q9BY89	Q9BY89 homo sapien
c 26	57.5	28.5	1511	4	Q75412	Q75412 homo sapien
c 27	57.5	28.5	1587	4	Q00508	Q00508 homo sapien
c 28	57	27.4	240	8	Q20933	Q20933 simuliun vi
c 29	57	28.2	302	13	Q319C3	Q319C3 anguilla an
c 30	57	28.2	880	10	Q9SVU3	Q9SVU3 arabisdopsi
c 31	57	28.2	924	10	Q949M4	Q949M4 arabisdopsi
c 32	57	28.2	3179	12	Q8V2A4	Q8V2A4 human herpe
c 33	56.5	27.2	203	6	Q29032	Q29032 sus scrofa
c 34	56.5	27.2	212	6	Q28512	Q28512 macaca mula
c 35	56.5	27.2	245	4	Q9NYC4	Q9NYC4 homo sapien
c 36	56.5	27.2	342	10	Q9W9M8	Q9W9M8 arabisdopsi
c 37	56	26.9	204	8	Q8SG48	Q8SG48 heteropsilo
c 38	56	26.9	315	16	P72593	P72593 synechocyst
c 39	56	26.9	363	4	Q9NY58	Q9NY58 homo sapien
c 40	56	27.7	468	11	Q9CYA6	Q9CYA6 mus musculu
c 41	56	27.7	468	11	Q31WQ1	Q31WQ1 mus musculu
c 42	56	27.7	511	4	Q96K62	Q96K62 homo sapien
c 43	56	27.7	554	5	Q9W4C1	Q9W4C1 drosophila
c 44	56	26.9	648	5	Q9NMD7	Q9NMD7 drosophila
c 45	56	26.9	701	5	Q9VJU4	Q9VJU4 drosophila

ALIGNMENTS

RESULT 1
ID Q80822 PRELIMINARY: PRT: 83 AA.

AC Q80822;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE Protein 11 xv.
GN XV.
OS Human T-cell leukemia virus type II (HTLV-II).
OC Viruses; Retrovird viruses; Retroviridae; Deltaretrovirus.
OX NCBITaxID=11909;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95297146; PubMed=7539968;
RA Clininale V., D'Agostino D.M., Zotti L., Franchini G., Felber B.K.,
RA Chieco-Bianchi L.;
RT "Expression and characterization of proteins produced by mRNAs spliced
RT into the X region of the human T-cell leukemia/lymphotropic virus type
RT II.";
RL Virology 209:445-456(1995).
DR EMBL: L1675; AAA98638.1;
SQ SEQUENCE 83 AA; 8427 MW; C4615BC04D9FD6B4 CRC64;

Alignment Scores:


```

FT  NON TER      210      210
SQ  SEQUENCE      210 AA; 23025 MW; 31119E4CA823633D CRC64;

Alignment Scores:
Pred. No.:          9.55          Length:          210
Score:              60.50         Matches:         17
Percent Similarity: 56.76%         Conservative:    4
Best Local Similarity: 45.95%       Mismatches:     11
Query Match:        29.09%         Indels:          5
DB:                  6             Gaps:           3

US-09-513-999C-3792_COPY_51_161 (1-111) x Q28273 (1-210)
OY  3 GGGTGCATCTTTTGGCTTCGACAGATCTTTTTCATCTTT-----GCAGGACTTCT 53
    |||||          |||||          |||||
Db  125 G1YrIpIeSerIeuRrIpYsIeSerHeIleWetherSerIaGIySerIu 144
OY  54 GGGGCGCGA---GTATGTAAACTCTGGCTCTGTGTGTGCTAGTG 101
    |||||          ::|||          |||::          |||::
Db  145 G1YlaIeYGIaIaIeUaIaSerProIySer---CysIeuGIuIaPhe 160

RESULT 6
O9N4W3
ID  O9N4W3          PRELIMINARY;          PRT:          644 AA.
AC  O9N4W3;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DI  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Y46H3C.4 protein.
CN  Y46H3C.4
OS  Caenorhabditis elegans.
OC  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC  Rhabditidae; Peloderinae; Caenorhabditis.
OY  NCBI_TaxID=6239;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2;
RC  MEDLINE=99069613; PubMed=9651916;
RA  None;
RT  "Genome sequence of the nematode C. elegans: a platform for
RT  investigating biology. The C. elegans Sequencing Consortium.";
RL  Science 282:2012-2018(1998).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2;
RA  Stoneking T., Wohlmann P., Lennox S.;
RT  "The sequence of C. elegans cosmid Y46H3C.";
RL  Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2;
RA  Waterston R.;
RL  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AC006776; AAF60626.1; -.
DR  HSSP; P06786; 1BGW.
DR  InterPro; IPR002205; DNA_topoisolv.
DR  Pfam; PF00521; DNA_topoisolv; 1.
DR  ProDom; PD000742; DNA_topoisolv; 1.
DR  SMART; SM00434; TOP4c; 1.
SQ  SEQUENCE 644 AA; 73898 MW; C530B213895D5958 CRC64;

Alignment Scores:
Pred. No.:          9.98          Length:          644
Score:              60.50         Matches:         17
Percent Similarity: 53.66%         Conservative:    5
Best Local Similarity: 41.46%       Mismatches:     14
Query Match:        29.95%         Indels:          5
DB:                  5             Gaps:           1

US-09-513-999C-3792_COPY_51_161 (1-111) x Q9N4W3 (1-644)
OY  108 AGAGCAGCCACTAGGCGCACACAGAGACCGACGAGTTTTCATCTCGGCCCCAGAG 49
    |||          |||          |||          |||          |||          |||

```

```

DB 281 ArgThrSerThrSerProGluValIleThrGlnIlePheTyrAspSerArgGlnGluLys 300
QY 48 TCCCGCAAGATGAAAAAGATCTCTG-----CAAGCGAAAAGATCCACC 4
DB 301 TTTLeuGlnArgLysGlnTyrLeuGlnValLeuGlnAlaGlnSerLysArgLeuThr 320
QY 3 CAT 1
DB 321 Asn 321

RESULT 7
Q20146 PRELIMINARY; PRT; 314 AA.
AC Q20146;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 36.0 kDa protein.
GN F38B6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA W11Cox L.;
RT "The sequence of C. elegans cosmid F38B6."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL EMBL; 040060; AAA81145.1; -.
DR InterPro: IPR002899; WRI/EB.
DR SMART; SM00289; WRI; 1.
KW Hypothetical protein.
SQ SEQUENCE 314 AA; 36010 MW; 6C5042D23E8038E6 CRC64;

Alignment Scores:
Pred. No.: 11.4 Length: 314
Score: 60.00 Matches: 11
Percent Similarity: 62.96% Conservative: 6
Best Local Similarity: 40.74% Mismatches: 6
Query Match: 28.85% Indels: 4
DB: 5 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x Q20146 (1-314)

QY 11 CTTTGGCTGACGAGATCTTTTCATCTT-----TGCAGGAGACTCTGGGCG 58
DB 278 TTTLeuLysGlnArgLysGlnTyrLeuGlnValLeuGlnAlaGlnSerLysArgLeuThr 297
QY 59 CGGAGTATGTAACACTGCTGG 79
DB 298 ArgLysLeuGlnLysAsnAsnTyr 304

RESULT 8
P97320 PRELIMINARY; PRT; 533 AA.
AC P97320;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)

```

```

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Testis-specific protein, DDC8.
GN DDC8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CD1;
RA Catalano R.D., Vlad M., Kennedy R.C.;
RT "Differential display to identify and isolate novel genes expressed
RT during spermatogenesis."
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y09878; CAA71005.1; -.
DR MGD; MG1:1929713; Ddc8.
SQ SEQUENCE 533 AA; 62039 MW; C7F75F456B24F52 CRC64;

Alignment Scores:
Pred. No.: 11.6 Length: 533
Score: 60.00 Matches: 11
Percent Similarity: 52.94% Conservative: 7
Best Local Similarity: 32.35% Mismatches: 16
Query Match: 29.70% Indels: 0
DB: 11 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x P97320 (1-533)

QY 107 GAGCAGCCACCTGACGACACAGAGCCAGAGTTTACATCTCGGCCGACAGAGT 48
DB 116 GtUGluArgGlyArgArgGlnGlnHisProLysSerArgLysLysAla 135
QY 47 CCTGCAAGATGAAAAAGATCTCGCAAGCGCAAAAGATCCA 6
DB 136 ProCysSerGluArgSerSerAlaLysAlaArgPro 149

RESULT 9
Q9LFLO PRELIMINARY; PRT; 1145 AA.
AC Q9LFLO;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cellulose synthase catalytic subunit-like protein.
GN F2K13.60.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project.
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391141; CAC01704.1; -.
DR InterPro: IPR005150; Cellulose_synth.
DR Pfam: PF03552; Cellulose_synth; 1.
SQ SEQUENCE 1145 AA; 128359 MW; 21A37FD050DC26BC CRC64;

Alignment Scores:
Pred. No.: 12 Length: 1145
Score: 60.00 Matches: 13
Percent Similarity: 57.14% Conservative: 3
Best Local Similarity: 46.43% Mismatches: 11
Query Match: 28.85% Indels: 1
DB: 10 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x Q9LFLO (1-1145)

```

```

DT      01-JAN-1998 (TREMBLrel_05, last sequence update)
DT      01-MAR-2002 (TREMBLrel_20, last annotation update)
DE      UNC-44 ankrylins.
GN      UNC-44.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC      Rhabdilitidae; Peloderrinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=N2;
RX      MEDLINE=95263663; Pubmed=7744957;
RA      Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,
RA      Boontarakulponlawee P., Jeyaprasath A., Hedgecock E., Wheaton V.I.,
RA      Sobery A.;
RT      "An ankrylin-related gene (unc-44) is necessary for proper axonal
RT      guidance in Caenorhabditis elegans.";
RL      J. Cell Biol. 129:1081-1092(1995).
RN      [2]
RP      REVISIONS, AND SEQUENCE OF 6126-6994 FROM N.A.
RC      STRAIN=N2;
RA      Otsuka A.J.;
RL      Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR      EMBL: U39847; AAB41827.1; -.
DR      EMBL: U21733; AAB384.1; -.
DR      HSSP; P42773; 1IHB.
DR      InterPro: IPR002110; ANK.
DR      InterPro: IPR000488; Death.
DR      InterPro: IPR001360; GH_1.
DR      InterPro: IPR002383; GLA_blood.
DR      InterPro: IPR000906; ZU5.
DR      Pfam; PF00023; ank; 24.
DR      Pfam; PF00531; death; 1.
DR      Pfam; PF00791; ZU5; 1.
DR      PRINTS; PRO1415; ANKYRIN.
DR      PRINTS; PRO0001; GLABLOOD.
DR      SMART; SM00248; ANK; 21.
DR      SMART; SM00005; DEATH; 1.
DR      SMART; SM00218; ZU5; 1.
DR      PROSITE; PS50088; ANK_REPEAT; 22.
DR      PROSITE; PS50297; ANK_REP_REGION; 1.
DR      PROSITE; PS50017; DEATH_DOMAIN; 1.
DR      PROSITE; PS00572; GLYCOSTYL_HYDROL_F1_1; UNKNOWN_2.
KW      ANK repeat; Repeat.
SQ      SEQUENCE 6994 AA; 775364 MW; 90CB449925D9923D CRC64;

Alignment Scores:
Pred. No.:          12 9           length:        6994
Score:              60.00         Matches:       11
Percent Similarity: 51.85%        Conservative:   3
Best Local Similarity: 40.74%     Mismatches:    13
Query Match:        29.70%        Indels:        0
DB:                  5            Gaps:          0

US-09-513-999C-3792_COPY_51_161 (1-111) x Q17343 (1-6994)
QY      107 GAGCAGCCACTCAGACGACACAGACCAGGATTTTACATTCTCGCCCCAGAAGT 48
      ||| ||||| ||| ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2796 GUGSERPROTIEGLNHSIGLVSGIVThrGIUGIVPhaeSPHisSerHisProGIUBer 2815
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      47 CCCTGC AAAAGATGA AAAAGAA 27
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2816 ProVallauSerGIUGlySGlu 2822

RESULT 12
Q9Y067 PRELIMINARY; PRT; 1167 AA.
AC      Q9Y067;
DT      01-NOV-1999 (TRMBLrel_12, Created)
DT      01-NOV-1999 (TRMBLrel_12, last sequence update)
DT      01-DEC-2001 (TRMBLrel_19, last annotation update)
DE      Tashatiz protein.
GN      TASHAT2.
```

OS *Theileria annulata*.
OC Eukaryota: Alveolata, Apicomplexa, Piroplasmida; Theileridae;
OC Theileriella.
OX NCBI_TaxId=5874;
OC [1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=9339393; PubMed=10413048;
RX Swan D.G., Phillips K., Tait A., Shields B.R.;
RA "Evidence for localisation of a Theileria parasite AT hook DNA-binding
RT Mol. Biochem. Parasitol. 101:117-129(1999).
RL EMBL: AJ132045; CAB42096.1; -
DR InterPro: IPR000637; AT.hook.
DR Pfam: PF02178; AT.hook. 3.
DR PRINTS: PR00529; ATHOOK.
DR SMART: SM00384; AT_hook. 3.
KW DNA-binding.
QC SEQUENCE 1167 AA; 132551 MW; 73EA7A6D6A32F91A CRC64;

Assignment Scores:		
Read, No.:	14.1	1167
Score:	59.50	16
Percent Similarity:	52.78	Conservative: 3
Best Local Similarity:	44.44	Mismatches: 15
Query Match:	29.46	Indels: 2
DB:	5	Gaps: 1

[illegible]

RESULT 13	
08UEA8	PRELIMINARY; PRT; 382 AA.
ID	08UEA8;
AC	01-JUN-2002 (TREMblrel, 21, Created)
DT	01-JUN-2002 (TREMblrel, 21, Last sequence update)
DT	01-JUN-2002 (TREMblrel, 21, Last annotation update)
DE	Hypothetical protein Atu1852.
GN	ATU1852 OR AGR_C_3398.
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
NC	Rhizobiaceae; Rhizobium.
NCBI_TaxID=	176299;

RP
RX
RR
RA
RA
RA
RA
RA
RA
RA
RA
RA
RA
RA
RA
RA
RA
RT
RT
RT
RL

SEQUENCE FROM N.A.
MEDLINE-21608550; PubMed-11743193;
Wood D.W., Setubal J.C., Kaul R., Monts D.E., Kitajima J.P.,
Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Ut., Woo L.,
Chen Y., Paulsen I.T., Eelsen J.A., Karp P.D., Bovee D. Sr.,
Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
Kutayavati T., Levy R., Li M.-J., McClelland E., Palmieri A.,
Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krepsan W., Ferry M.,
Gordon-Kamm B., Liao L., Kim S., Hendick C., Zhao Z.-Y., Dolan M.,
Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
Nester E.W.;
C58 "": The genome of the natural genetic engineer *Agrobacterium tumefaciens*
Science 294:2317-2323(2001).

RA MEDLINE=21608551, PubMed=11743194; Miller N, Blanchard M,
RA Goodner B, Hinkle G, Gattung S, Askenazi M, Halling C, Mullin L,
RA Qurello B, Goldman B.S., Cao Y, Askew M, Halling C, Mullin L,
RA Hounell K, Gordon J., Vaudin M, Iartchouk O, Epp A, Liu F,
RA Wollam C, Allinger M., Doughy D., Scott C., Lapps C., Markelz B,
RP SEQUENCE FROM N.A.

RA Riancho C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.
RA Clano C., Slater S.
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:323-328(2001).
RL EMBL, AE009139, AA142648.1, -
DR EMBL, AE008105, AA67619.1, -
KW Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 362 AA; 42256 MW; 21ACD2A21C7577F CRC64;

Alignment Scores:	
Pred. No.:	15.8
Score:	59.00
Percent Similarity:	43.90%
Best Local Similarity:	36.59%
Query Match:	28.37%
DB:	16
Length:	38
Matches:	3
Conservative:	1
Mismatches:	1
Indels:	2
Gaps:	2

US-09-513-999C-3792_COPY_51_161 (1-111) x Q8UEA8 (1-382)

QY	6	TGATCTTTTGGCTGACGGATTTCTTTTCACTTTT	-----GCA	44
Db	116	TyrArgPhe-----GlyLeuPhePheIaPheProPheLeuMetLeuIaGly	132	
QY	45	GGACATCTGGGGCCGAGTGTGTAACCCCGGGATCTGTGTGTCCTCGAGTGGCTG	104	
Db	133	GlyAlaLeuAlaAlaGlyIlyAlaLeuSerProTyrIleuPheSerLeuProIeuTyrIleu	155	

QY	105	CTC	107
		111	
Db	153	Leu	153

RESULT 14	
Q9QYS6	PRELIMINARY;
ID Q9QYS6	PRT; 537 AA

DE Soss2 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxId=10090;
RN [1]
RP SPOUNGE FROM N.A.
CP EMPAIN-126/0074

RA	Esteban L.M., Lengal C., Santos E.;
RT	"Evolutionary conservation of genomic structure among different mammalian GEF genes.";
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBD databases.
DR	EMBL; AF094685; AAF08010.1; -; JOINED.
DR	EMBL; AF094685; AAF08010.1; JOINED.
DR	EMBL; AF094681; AAF08010.1; JOINED.
DR	EMBL; AF094682; AAF08010.1; JOINED.
DR	EMBL; AF094683; AAF08010.1; JOINED.
DR	EMBL; AF094684; AAF08010.1; JOINED.
DR	InterPro: IPR002965; P_tch_extensn.
DR	InterPro: IPR001895; RasGEF_CDC25.
DR	Pfam: PF00617; RasGEF. 1.
DR	PRINTS: PRO1217; PRICHEXTENSXN.
DR	SMART: SMO0147; RasGEF. 1
DR	PROSITE; PS00720; GDS_CDCC25; 1.
PT	NON_TER
Q	SEQUENCE 537 AA; 60255 MW; 578B5A7BD0D7D1B4 CRC64;

Alignment Scores:	
Pred. No.:	16, 1
Score:	59.00
Percent Similarity:	48.57%
Best Local Similarity:	37.14%
Query Match:	29.21%
DB:	11
Length:	537
Matches:	13
Conservative:	4
Mismatches:	12
Indels:	6
Gaps:	1

US-09-513-999c-3792_COPY_51_161 (1-111) x 090XS6 (1-537)

QY 106 AGCAGCCACTTCAGCAGCAGCAGCAGCAGGAGTTTACATCTCCGGCCCCAGAGTC 47
 Db 484 SerSerHisSer-----SerLeuAlaHisLeuProAlaProProVal 497

QY 46 CCTGCAAAAGATGAAAAAGATCTGCAGAGCAAGATCATCCACCA 2
 Db 498 ProProArgGlnAsnSerSerProLeuProLysLeuProPro 512

RESULT 15

Q63122 PRELIMINARY; PRT; 230 AA.
 ID 063122;
 AC 063122;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Alpha-3 type IV collagen (Fragment).
 IN COL4A3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY CORTEX;
 RX MEDLINE=98210005; PubMed=9550634;
 RA Ryan J.J., Katbama I., Mason P.J., Pusey C.D., Turner A.N.;
 RT "Sequence analysis of the 'Goodpasture antigen' of mammals."
 RL Nephrol. Dial. Transplant. 13:602-607(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY CORTEX;
 RA Turner N.;
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: I47281; AAB72238.2;
 DR InterPro: IPR001442; ProcollagenC4.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF01413; C4; 2.
 DR ProDom: PD003923; ProcollagenC4; 1.
 DR SMART: SM00111; C4; 2.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
 KW Collagen.
 FT NON_TER 1
 FT NON_TER 230
 SQ SEQUENCE 230 AA; 25398 MW; 29549E5314CC056 CRC64;

Alignment Scores:

Ref. No.:	18.2	Length:	230
Score:	58.50	Matches:	15
Percent Similarity:	56.76%	Conservative:	6
Best Local Similarity:	40.54%	Mismatches:	11
Query Match:	28.12%	Indels:	5
Db:	11	Gaps:	3

US-09-513-999c-3792_COPY_51_161 (1-111) x Q63122 (1-230)

QY 3 GGGTGGATCTTTGCTTCAGAGATCTTTTCATCTT-----GCAGGCACTTCT 53
 Db 133 GYTTPVALSerLeuTyrLysGlyPheSerPheValMetPheThrSerAlaGlySerGlu 152
 QY 54 GGGGCCGGA--GATGTAAACCTCTGGGCTCTGTGTGTGCTGCTGAGTGG 101
 Db 153 GYAlaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGluGluPhe 168

Search completed: April 30, 2003, 14:36:15
 Job time : 32.5 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 30, 2003, 14:34:42 ; Search time 9 Seconds
(without alignments)
1023.083 Million cell updates/sec

```
Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 208
Sequence: 1 atggtgatctctttgcctt.....gctgagtgctgctctact 111
```

Scoring table:		
BLOSUM62		
Xgapop 10.0 ,	Xgapext 0.5	
Ygapop 10.0 ,	Ygapext 0.5	
Fgapop 6.0 ,	Fgapext 7.0	
Delop 6.0 ,	Delext 7.0	

searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

Command line parameters:
-MODEL=frimmet_n2p.model -DEV=xip
-O=/cgen2.1/USPPO-SWIFT/fasta1513999/runat.30042003.143401.25593/app_query.fasta.1.2633
-DB=fastaSport.40 -QEMT=fasta1513999.SUPPLY-n2p.rsp -MISMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blissom62 -TRANS=human40.cdi
-LIST=43 -DOCALLIG=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US059151399.qcgen.1.1.26_etunal.30042003.143401.25593 -ICPU=3
-NO_XLPPY -NO_MAXIP -LARGEJOINTS -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Length	DB	ID	Description
C	1	63.5	31.4	648	1	Z202_HUMAN	Q95125 homo sapien
	2	60.5	29.1	471	1	CA34_BOVIN	Q28084 bos taurus
	3	59	29.2	1297	1	S2382_MOUSE	Q02384 mus musculus
	4	58.5	29.0	668	1	SC6L_HUMAN	Q05171 homo sapien
	5	58	27.9	391	1	CATE_CAVPO	P25796 cavia porcea
	6	57	28.2	3149	1	TEGUE_EBV	P03186 epstein-barr
	7	56.5	28.0	938	1	CAP4_HUMAN	Q13111 homo sapien
	8	56.5	27.2	1670	1	CA34_HUMAN	Q01935 homo sapien
	9	56	26.9	396	1	CATE_HUMAN	P14091 homo sapien
	10	55.5	26.7	147	1	TTYH_MOUSE	P07309 mus musculus
	11	55.5	26.7	687	1	VS41_GILDA	P92127 giardia lam
	12	54.5	27.0	113	1	N12B_MESEA	Q40339 medicago sa
	13	54.5	26.2	150	1	TTYHV_PIG	P50390 sus scrofa
	14	54.5	26.2	397	1	UT2_RAT	Q62668 rattus norv
	15	54.5	26.2	481	1	WGLJ_JUVIN	P26313 junco arena
	16	54.5	27.0	1520	1	TOP2_CABEL	Q23670 caenorhabdit
	17	54	26.0	387	1	CYB_PODAN	P02659a podospira a
	18	54	26.7	731	1	MR11_COCCI	Q94919 coprinus ci

C	19	53.5	26.5	63.4	1	HNFA_CHICK
C	20	53.5	26.5	98.6	1	AC1S_DROME
C	21	53	26.2	118	1	VE4_HPV13
C	22	53	26.2	346	1	GCP_BOBBU
C	23	53	25.5	397	1	CATE_MOUSE
C	24	53	25.5	398	1	CATE_RAT
C	25	53	26.2	716	1	PEP_DROME
C	26	52.5	25.2	147	1	TTHY_BOVIN
C	27	52.5	25.2	389	1	UT1_HUMAN
C	28	52.5	25.2	531	1	FNX1_SCHPO
C	29	52.5	26.0	768	1	CNIC_RAT
C	30	52	25.0	204	1	GTR5_BOVIN
C	31	52	25.0	587	1	LACP_KUUA
C	32	52	25.7	1581	1	LMG3_MOUSE
C	33	52	25.7	1859	1	GBF1_HUMAN
C	34	51.5	24.8	147	1	TTHY_RAT
C	35	51.5	25.5	248	1	DJ_DROME
C	36	51	25.2	50	1	RS30_AERPE
C	37	51	25.2	269	1	HXB5_HUMAN
C	38	51	25.2	269	1	HXB5_MOUSE
C	39	51	25.2	318	1	KHXB_VIBCH
C	40	51	24.5	354	1	NOV_MOUSE
C	41	51	24.5	452	1	AD1L_XENIA
C	42	51	25.2	432	1	CAT3_SOYBN
C	43	51	24.5	679	1	SYGB_BACSU
C	44	51	24.5	866	1	NASA_KLEPN
C	45	51	25.2	1119	1	RPOB_TJHQA

ALIGNMENTS

RESULT 1			
ID	NAME	STANDARD;	PRT; 648 AA.
AC	Z202_HUMAN		
DT	09J125; Q9H1B9;		
DT	15-JUL-1999 (Rel. 38, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	zinc finger protein 202.		
GN	ZNF202.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT VAL-154.		
RC	TISSUE=Testis;		
RX	MEDLINE=99009320; PubMed=9799754;		
RA	Monaco C., Helmer Citterich M., Caprini E., Vorechovsky I., Russo G.,		
RT	Croce C.M., Barbanti-Brodano G., Negri M.;		
RT	"Molecular cloning and characterization of ZNF202: A new gene at		
RL	11q23.3 encoding testis-specific zinc finger proteins.";		
RL	Genomics 52:358-362(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Langmann T., Potsch-Ozcurumez M., Helmerl S., Andrikovics H.,		
RT	Schmitt J.;		
RT	"Genomic sequence analysis of the ZNF202 gene: relevance for lipid		
RT	parameters.";		
RL	submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	CHARACTERIZATION.		
RX	MEDLINE=20283591; PubMed=10748193;		
RA	Wagner S., Hess M.A., Ormonde-Hanson P., Malandro J., Hu H., Chen M.,		
RA	Kelner R., Frodsham M., Schumacher C., Beluch M., Honer C.,		
RA	Skolnick M., Ballinger D., Bowen B.R.;		
RT	"A broad role for the zinc finger protein ZNF202 in human lipid		
RT	metabolism.";		
RL	J. Biol. Chem. 275:15685-15690(2000).		
CC	-I- FUNCTION: TRANSCRIPTIONAL REPRESSOR THAT BINDS TO ELEMENTS FOUND		
CC	PREDOMINANTLY IN GENES THAT PARTICIPATE IN LIPID METABOLISM. AMON		
CC	ITS TARGETS ARE STRUCTURAL COMPONENTS OF LIPOPROTEIN PARTICLES		
CC	(APOLOPROTEINS AIV, CIIL, AND E). ENZYMES INVOLVED IN LIPID		

```

Query Match: 31.44% Indels: 5
DB: 1 Gaps: 3

US-09-513-999C-37992_COPY_51_161 (1-111) x 2202_HUMAN (1-648)
Oy 110 GTAGAGCAGCCGACGACACACAGAGCCAGG-----AGTTTACATACGCCGC 57
      ::::::::::: :::: ||||| ||| ||||| ||||| ::::: |||
Db 300 IlegInGUPro---GInGUthGInGUProGUlIUleuSerPherThyTrmGly 318
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 56 CCCAGAGTCCCTGCAGATGAAAGAAATCTCTGCAAGCCAAAGAT 9
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 319 AspArgSer-----LysAspGInGUlUGlucSerIeUGInGUlUasp 332
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
CA34_BOVIN STANDARD; PRT; 471 AA.
ID CA34_BOVIN
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Collagen alpha 3(IV) chain (Fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093146; PubMed=1985905;
RT Morrison K.E., Germino G.G., Reeders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
RT encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RT Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RC TISSUE=Kidney;
RX MEDLINE=8830844; PubMed=3417661;
RT Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
RT of collagen IV.";
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RC TISSUE=Kidney;
RX MEDLINE=87222419; PubMed=2438283;
RT Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RA "Localization of the Goodpasture epitope to a novel chain of basement
RA membrane collagen.";
RT RT
RT J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC MEMBRANE BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND EXTRACIN/
CC NIDOEEN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

```

```
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M63139; AAA62708.1; -.
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001442; ProcollagNC4.
CC Pfam: PF01391; Collagen; 4.
CC DR Pfam: PF01413; C4; 2.
CC DR ProDom: PD003923; ProcollagNC4; 2.
CC DR SMART: SM00111; C4; 2.
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Basement membrane; Collagen; Cell adhesion.
CC FT NON_TER 1 238 TRIPLE-HELICAL REGION.
CC FT DOMAIN 239 471 NONHELICAL REGION (NC1).
CC FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
CC FT MOD_RES 232 232 HYDROXYLATION.
CC FT MOD_RES 238 238 HYDROXYLATION.
CC FT DISULFID 261 352 OR 349 (BY SIMILARITY).
CC FT DISULFID 294 349 OR 352 (BY SIMILARITY).
CC FT DISULFID 306 312 BY SIMILARITY.
CC FT DISULFID 371 466 OR 463 (BY SIMILARITY).
CC FT DISULFID 405 463 OR 466 (BY SIMILARITY).
CC FT DISULFID 417 423 BY SIMILARITY.
CC FT CONFLICT 253 253 S -> Y (IN REF. 3).
CC SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Alignment Scores:
Pred. No.: 5.2 Length: 471
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 1 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x CA34_BOVIN (1-471)
QY 3 GGGTGGATCTTTTGCCTTGACGAGATCTTTTCATCTTT-----GCAGGACTTCT 53
DB 374 GYTPRIleSerleuTrpIysGlyPheSerPheIleMetPheThrSerIaGlySerGlu 393
QY 54 GGGGCGCGGA---GTATGTAAACTCCTGGCTCTGTGTGTGCGCTGAGTGG 101
DB 394 GYIaGlyGlyGlnaIaLeuAlaSerProGlySer---CysleuGlnGluPhe 409

RESULT 3
SOS2_MOUSE
ID SOS2_MOUSE STANDARD: PRT; 1297 AA.
AC 002384;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Son of sevenless protein homolog 2 (SOS-2) (mSOS-2) (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Eye;
RX MEDLINE=92335328; PubMed=1631150;
```

```
RA Bowtell D.D., Fu P., Simon M.A., Senior P.V.;
RT "Identification of murine homologues of the Drosophila son of
RT sevenless gene: potential activators of ras.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST EMBRYONIC AND ADULT TISSUES.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOG (DH) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11664; CAA77732.1; -.
CC PIR: S25714; S25714.
CC DR HSSP: O62245; IPMS.
CC DR MGD: MGI:98355; Sos2.
CC DR InterPro: IPR001331; GDS_CDC24.
CC DR InterPro: IPR001849; PH.
CC DR InterPro: IPR000651; RasGEFN.
CC DR InterPro: IPR001895; RasGEF_CDC25.
CC DR InterPro: IPR000219; RhogEF.
CC DR Pfam: PF00169; PH; 1.
CC DR Pfam: PF00617; RasGEF; 1.
CC DR Pfam: PF00618; RasGEFN; 1.
CC DR Pfam: PF00621; RhogEF; 1.
CC DR SMART: SM00233; PH; 1.
CC DR SMART: SM00147; RasGEF; 1.
CC DR SMART: SM00229; RasGEFN; 1.
CC DR SMART: SM00325; RhogEF; 1.
CC DR PROSITE: PS00741; DH_1; FALSE_NEG.
CC DR PROSITE: PS00720; DH_2; 1.
CC DR PROSITE: PS00720; GDS_CDC25; 1.
CC DR PROSITE: PS50003; PH_DOMAIN; 1.
CC KW Guanine-nucleotide releasing factor.
CC FT NON_TER 1 1
CC FT DOMAIN 164 353 DH.
CC FT DOMAIN 407 510 PH.
CC FT DOMAIN 740 926 RAS-GEF.
CC FT DOMAIN 1145 1148 POLY-PRO.
CC FT DOMAIN 1164 1173 POLY-PRO.
CC SQ SEQUENCE 1297 AA; 148846 MW; 3A53F8515BF9D625 CRC64;

Alignment Scores:
Pred. No.: 8.17 Length: 1297
Score: 59.00 Matches: 13
Percent Similarity: 48.57% Conservative: 4
Best Local Similarity: 37.14% Mismatches: 12
Query Match: 29.21% Indels: 6
DB: 1 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x SOS2_MOUSE (1-1297)
QY 106 AGCAGCCACTCAGGACACAGAGACCCAGAGTTTACATCTCCGCGCCCAAGAGTC 47
DB 1244 SerSerHisSer-----SerIleuAlaHisIleuProAlaProProVal 1257
QY 46 CCTGCAAGATGAAAAAAGATCTCTGCAAGCAAAAGATCCACCA 2
DB 1258 ProProArgGlnAAsnSerProIleuProIysIleuProPro 1272

RESULT 4
SCEL_HUMAN
ID SCEL_HUMAN STANDARD: PRT; 668 AA.
AC 095171;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Scellin.
 GN SCEL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 443-457 AND 635-650.
 RC TISSUE=foreskin;
 RX MEDLINE=99030435; PubMed=9813070;
 RA Champliand M.-F., Burgess R.E., Jin W., Baden H.P., Olson P.F.;
 RT "CDNA cloning and characterization of scellin, a LIM domain protein
 of the keratinocyte cornified envelope.";
 RL J. Biol. Chem. 273:31547-31554(1998).
 CC -1- FUNCTION: May function in the assembly or regulation of proteins
 in the cornified envelope. The LIM domain may be involved in
 homotypic or heterotypic associations and may function to localize
 scellin to the cornified envelope.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. May become cross-linked to
 membrane proteins by transglutaminase.
 CC -1- TISSUE SPECIFICITY: Highly expressed in esophagus. It is also
 expressed in keratinocytes, amniotic tissue, foreskin stratum
 spinosum and stratum granulosum, hair follicle and nail.
 CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
 IONS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 DR EMBL: AF045941; AAC78461.1; -;
 DR Genbank: HGNC:10573; SCEL.
 DR MIM: 604112; -;
 DR InterPro: IPR001781; LIM.
 DR ProDom: PD000094; LIM; 1.
 DR SMART: SM00132; LIM; 1.
 DR PROSITE: PS00478; LIM_DOMAIN_1; FALSE_NEG.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
 KW LIM domain; Metal-binding; Zinc; Repeat.
 FT DOMAIN 599 665 16 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 231 543 1.
 FT REPEAT 231 246 1.
 FT REPEAT 247 266 2.
 FT REPEAT 267 286 3.
 FT REPEAT 287 306 4.
 FT REPEAT 307 326 5.
 FT REPEAT 327 346 6.
 FT REPEAT 347 366 7.
 FT REPEAT 367 386 8.
 FT REPEAT 387 406 9.
 FT REPEAT 407 426 10.
 FT REPEAT 427 445 11.
 FT REPEAT 446 464 12.
 FT REPEAT 465 484 13.
 FT REPEAT 485 503 14.
 FT REPEAT 504 523 15.
 FT REPEAT 524 543 16.
 SO SEQUENCE 668 AA; 75297 MW; 612BC3686DBA9FB0 CRC64;
 Alignment Scores:
 Pred. No.: 9 48
 Score: 58.50
 Percent Similarity: 56.108
 Best Local Similarity: 39.028
 Query Match: 28.968
 DB: 1
 Gaps: 3
 US-09-513-999c-3792_COPY_51_161 (1-111) x SCEL_HUMAN (1-668)

QY 106 AGCAGCCATCTCAGGACACA---CACAGAGACCCAGGAGTTTACATCTCCGCCCCAGAA 50
 DB 174 SerSerSerThrGlyThrArgArgGluProGlyVal---HisProProIleProPro 192
 QY 49 GTCCCGCAAGAAATGAAAAAAGATCCTGCAGAGS-----CAAAAGATCCAC 5
 DB 193 LysProSerSerProValSerSerProValSerProValGlnLeuArgGlnAspAsnArgGlnIleHis 212
 QY 4 CCA 2
 DB 213 Pro 213
 RESULT 5
 ID CATE_CAVPO STANDARD; PRT; 391 AA.
 AC P25796;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin E precursor (EC 3.4.23.34).
 GN CTSE.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=92355614; PubMed=1644829;
 RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Kolwal O.,
 RA Tanji M., Yakabe E., Althaus S.B., Takahashi K.;
 RT "Isolation, characterization, and structure of procathepsin E and
 cathepsin E from the gastric mucosa of guinea pig.";
 RT Adv. Exp. Med. Biol. 362:211-221(1995).
 CC -1- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
 LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
 specificity.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 DR EMBL: M88653; AAA37052.1; -;
 DR EMBL: S80547; AAB35844.1; -;
 DR PIR: A43356; A43356.
 DR HSP: P00794; 4CMS.
 DR MEROPS: A01.010; -;
 DR InterPro: IPR001461; Asproteaseal.
 DR InterPro: IPR001969; Asprotease_site.
 DR Pfam: PF00026; asp; 1.
 DR PRINTS: PR00792; PEPSIN.
 DR PROSITE: PS00141; ASP_PROTEASE; 2.
 KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 18
 FT PROPEP 19 53
 FT CHAIN 54 391
 CATEPSIN E.

FT ACT SITE 92 92 BY SIMILARITY.
 FT ACT SITE 276 276 BY SIMILARITY.
 FT DISULFID 56 56 INTERCHAIN (PROBABLE).
 FT DISULFID 105 110 BY SIMILARITY.
 FT DISULFID 267 271 BY SIMILARITY.
 FT DISULFID 309 346 BY SIMILARITY.
 FT CARBOHD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 391 AA; 42132 MW; 78D216BF8CFCDAD CRC64;

Alignment Scores:

Pred. No.: 11 Length: 391
 Score: 58.00 Matches: 12
 Percent Similarity: 63.33% Conservative: 7
 Best Local Similarity: 40.00% Mismatches: 11
 Query Match: 27.88% Indels: 0
 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x CAFE_CAVPO (1-391)

QY 1 ATGGGTGATCTTTTCCCTGAGAGATCTTTTCATCTTTCAGAGGAGCTTGGGCGC 60
 Db 127 ValGlyAsnSerPheSerIleGltIrycIythrGlySerIeuThrgIleIleGlyAla 146
 QY 61 GAGTATGTAACCTCTGCTCTGCTG 90
 Db 147 AspGlnValSerValGluGlyLeuThrVal 156

RESULT 6

TEGU_EBV STANDARD; PRT: 3149 AA.
 AC P03186:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Large tegument protein.
 GN BPLF1.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=103777;

RN SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tuffnell P.S., Barrell B.G.;
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RL Nature 310:207-211(1984).

CC -1- FUNCTION: TEGUMENT PROTEIN.

CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
 CC EBV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: V01555; CAA24839.1; .
 DR PIR: A03747; OQBR8.
 DR PIR: S32993; S32993.
 SQ SEQUENCE 3149 AA; 337954 MW; 3DD0C576587313D8 CRC64;

Alignment Scores:

Pred. No.: 14.9 Length: 3149
 Score: 57.00 Matches: 11
 Percent Similarity: 41.18% Conservative: 20
 Best Local Similarity: 32.35% Mismatches: 0
 Query Match: 28.22% Indels: 0
 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x TEGU_EBV (1-3149)

QY 103 AGCCACTAGGCACACACAGAGACCGAGATTTCATATCTCGGCCCAAGACTCCT 44
 Db 449 ThrHisValProProHisArgProProSerAlaAlaArgLeuProProValIlePro 468
 QY 43 GCAAAGATGAAAAGATCTCGCAAGCAAGCAAAAGATCCACCA 2
 Db 469 IleProHisGlnSerProProAlaSerProThrProHisPro 482

RESULT 7

CAFA_HUMAN STANDARD; PRT: 938 AA.
 ID Q13111; Q9UY8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chromatin assembly factor 1 subunit A (CAF-1 subunit A) (Chromatin
 DE assembly factor 1 p150 subunit) (CAF-1 150 kDa subunit) (CAF-Ip150).
 GN CHAF1A OR CAF1p150 OR CAF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;

RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Cervical adenocarcinoma;
 RX MEDLINE=95323966; PubMed=7600578;
 RA Kaufman P.D., Kobayashi R., Kessler N., Stillman B.;
 RT "The p150 and p60 subunits of chromatin assembly factor I: a molecular
 RT link between newly synthesized histones and DNA replication.";
 RL Cell 81:1105-1114(1995).

RN SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RX MEDLINE=21147928; PubMed=11250073;
 RA Dong H., Lin W., Zhang C.-K., Xiong H., Fu G., Jin W.-R., Chen R.,
 RA Chen Z., Qi Z.-T., Huang G.M.;
 RT "Genomic sequence and expression analyses of human chromatin assembly
 RT factor 1 p150 gene.";
 RL Gene 264:187-196(2001).

RN INTERACTION WITH PCNA, AND SUBCELLULAR LOCATION.
 RP MEDLINE=99159824; PubMed=10052459;
 RA Shibahara K., Stillman B.;
 RT "Replication-dependent marking of DNA by PCNA facilitates
 RT CAF-1-coupled inheritance of chromatin.";
 RL Cell 96:575-585(1999).

RN INTERACTION WITH PCNA.
 RP MEDLINE=20115866; PubMed=10648606;
 RA Moggs J.G., Grandt P., Quijvy J.P., Jonsson Z.O., Hubscher U.,
 RA Becker P.B., Almouzni G.;
 RT "A CAF-1-PCNA-mediated chromatin assembly pathway triggered by sensing
 RT DNA damage.";
 RL Mol. Cell. Biol. 20:1206-1218(2000).

RN REVIEW.
 RP MEDLINE=20351441; PubMed=10893180;
 RA Ridgway P., Almouzni G.;
 RT "CAF-1 and the inheritance of chromatin states: at the crossroads of
 RT DNA replication and repair.";
 RL J. Cell Sci. 113:2647-2658(2000).

CC -1- FUNCTION: COMPLEX THAT IS THOUGHT TO MEDIATE CHROMATIN ASSEMBLY IN
 CC DNA REPLICATION AND DNA REPAIR. ASSEMBLES HISTONE OCTAMERS ONTO
 CC REPLICATING DNA IN VITRO. CAF-1 PERFORMS THE FIRST STEP OF THE
 CC NUCLEOSOME ASSEMBLY PROCESS, BRINGING NEWLY SYNTHESIZED HISTONES
 CC H3 AND H4 TO REPLICATING DNA; HISTONES H2A/H2B CAN BIND TO THIS
 CC CHROMATIN PRECURSOR SUBSEQUENT TO DNA REPLICATION TO COMPLETE THE
 CC HISTONE OCTAMER. P150 BINDS TO HISTONES H3 AND H4. IT MAY PLAY A
 CC ROLE IN HETEROCHROMATIN MAINTENANCE IN PROLIFERATING CELLS BY
 CC BRINGING NEWLY SYNTHESIZED CBX PROTEINS TO HETEROCHROMATIC DNA
 CC REPLICATING FOCI (BY SIMILARITY).
 CC -1- SUBUNIT: CAF-1 IS COMPOSED OF THREE SUBUNITS, P48, P60 AND P150.

CC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=89380302; PubMed=26744141;
 RT Azuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;
 RT "Human gastric cathepsin E. Predicted sequence, localization to
 RT chromosome 1, and sequence homology with other aspartic
 RT proteinases.";
 RL J. Biol. Chem. 264:16748-16753(1989).
 [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92112877; PubMed=1370478;
 RT Azuma T., Liu W.G., Vander laan D.J., Howcock A.M., Taggart R.T.;
 RT "Human gastric cathepsin E gene. Multiple transcripts result from
 RT alternative polyadenylation of the primary transcripts of a single
 RT gene locus at 1q31-q32.";
 RL J. Biol. Chem. 267:1609-1614(1992).
 [3]
 RP SEQUENCE FROM N.A.
 RA Tanneil P.J., Kay J.;
 RT "Human procathepsin E.";
 RT Submitted (Oct-1999) to the EMBL/Genbank/DBJ databases.
 [4]
 RP SEQUENCE OF 54-68; 77-95; 141-155; 275-285 AND 389-396.
 RA MEDLINE=90241267; PubMed=2334440;
 RT Athauda S.B.P., Matsuzaki O., Kageyama T., Takahashi K.;
 RT "Structural evidence for two isozymic forms and the carbohydrate
 RT attachment site of human gastric cathepsin E.";
 RL Biochem. Biophys. Res. Commun. 168:878-885(1990).
 CC -1- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
 CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
 CC specificity.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M84424; AAA52300.1; -;
 DR EMBL: M84413; AAA52300.1; JOINED.
 DR EMBL: M84417; AAA52300.1; JOINED.
 DR EMBL: M84418; AAA52300.1; JOINED.
 DR EMBL: M84419; AAA52300.1; JOINED.
 DR EMBL: M84420; AAA52300.1; JOINED.
 DR EMBL: M84421; AAA52300.1; JOINED.
 DR EMBL: M84422; AAA52300.1; JOINED.
 DR EMBL: J05036; AAA52300.1; -;
 DR EMBL: AJ250717; CAB82850.1; -;
 DR PIR: A34401; A34401.
 DR PIR: A34643; A34643.
 DR PIR: A42038; A42038.
 DR HSSP: P00794; ACMS.
 DR HSSP: A01_010; -;
 DR Genew: HGNC:2530; CTSE.
 DR MIM: 116890; -;
 DR InterPro: IPR001461; AsparticaseA1.
 DR InterPro: IPR001969; Asparticase_sib.
 DR Pfam: PF00026; asp_1.
 DR PRINTS: PR00792; PEPsin.
 DR PROSITE: PS00141; ASP_PROTEASE; 2.
 KW hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 53
 FT CHAIN 54 396
 FT MOD_RSS 18 18
 FT ACT_SITE 96 96
 FT BY SIMILARITY.

FT ACT_SITE 281 281 BY SIMILARITY.
 FT DISULFID 60 60 INTERCHAIN (PROBABLE).
 FT DISULFID 109 114 BY SIMILARITY.
 FT DISULFID 272 276 BY SIMILARITY.
 FT DISULFID 314 351 BY SIMILARITY.
 FT CARBOHYD 90 90 N-LINKED (GLCNAC...).
 FT CARBOHYD 220 220 O-LINKED (POTENTIAL).
 FT CARBOHYD 333 333 O-LINKED (POTENTIAL).
 SQ SEQUENCE 396 AA; 42793 MW; 40B64365FB01521E CRC64;
 Alignment Scores:
 Pred. No.: 20 Length: 396
 Score: 56.00 Matches: 12
 Percent Similarity: 62.07% Conservative: 6
 Best Local Similarity: 41.38% Mismatches: 11
 Query Match: 26.92% Indels: 0
 DB: 1 Gaps: 0
 US-09-513-999c-3792_COPY_51_161 (1-111) x CATE_HUMAN (1-396)
 QY 4 GGTCGATCTTTGGCTTCAGAGATCTTTTCATCTTGCAGGACTTTCGGCCGAG 63
 Db 132 GlnGlnSerPheSerIleGlnIleGlyThrGlySerLeuSerGlyIleIleGlyAlaAsp 151
 QY 64 TATGTAAACCTCGGGCTCTGTCTG 90
 Db 152 GlnValSerValGlnGlyLeuThrVal 160
 RESULT 10
 ID THY_MOUSE STANDARD: PRT: 147 AA.
 AC P07309;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transferrin precursor (Prelalbumin).
 GN TTR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86139970; PubMed=3005251;
 RT Wakisugi S., Maeda S., Shimada K., Nakashima H., Migita S.;
 RT "Structural comparisons between mouse and human prealbumin.";
 RL J. Biochem. 98:1707-1714(1985).
 [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=87008480; PubMed=3020014;
 RT Wakisugi S., Maeda S., Shimada K.;
 RT "Structure and expression of the mouse prealbumin gene.";
 RL J. Biochem. 100:49-58(1986).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Choroid plexus;
 RA Kita H., Kawamoto S., Okubo K., Matsubara K.;
 RT Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: Thyroid hormone-binding protein. Probably transports
 CC thyroxine from the bloodstream to the brain.
 CC -1- SUBUNIT: Homotrimer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN THE CHOROID PLEXUS. ALSO
 CC PRESENT IN THE LIVER.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----


```

CC -----
DR EMBL: X03351; CAA27057.1; -
DR EMBL: D00073; BAA0050.1; -
DR EMBL: D00071; BAA0050.1; JOINED.
DR EMBL: D00072; BAA0050.1; JOINED.
DR EMBL: D89076; BAA13757.1; -
DR PIR: A24132; VBMS.
DR HSSP: P02767; IGKE.
DR MGD; MGI:98865; TLR.
DR InterPro: IPR000895; Transthyretin.
DR Pfam: PF00576; Transthyretin.1.
DR PRINTS: PR00189; TRNSTHYRETIN.
DR ProDom: PD003457; Transthyretin.1.
DR SMART; SM00095; TR_THY.1.
DR PROSITE; PS00768; TRANSTHYRETIN_1; 1.
DR PROSITE; PS00769; TRANSTHYRETIN_2; 1.
DR Albumin; Transport; Retinol-binding; Vitamin A; Thyroid hormone;
KW Signal.
FT SIGNAL 1 20
FT CHAIN 1 147 TRANSTHYRETIN
FT BINDING 35 35 THYROID HORMONES (BY SIMILARITY).
FT BINDING 74 74 THYROID HORMONES (BY SIMILARITY).
SQ SEQUENCE 147 AA; 15776 MW; 9803CCC3024BA911 CRC64;

Alignment Scores:
Pred. No.: 23.2 Length: 147
Score: 55.50 Matches: 12
Percent Similarity: 64.00% Conservatve: 4
Best Local Similarity: 48.00% Mismatches: 8
Query Match: 26.68% Indels: 1
DB: 1 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x TTHY_MOUSE (1-147)

QY 6 TGGATCTTTTGGCTTGCAGAGATCTTTTCATCTTT---GCAGGAGCTTGGGGCCGGA 62
Db 7 PheLeuLeuCySLeuAlaGlyLeuValPheValSerGluAlaGlyProAlaGlyAlaGly 26
QY 63 GATGTAACACTGCT 77
Db 27 GluSerLysCysPro 31

RESULT 11
VS41_GIALA
ID VS41_GIALA STANDARD: PRT; 687 AA.
AC P92127;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Variant-specific surface protein VSP41 precursor (CRISP-90).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
OX NCBI_TaxID=5741;
RN RP SEQUENCE FROM N.A.
RC STRAIN=O2-4A1;
RX MEDLINE=97321554; PubMed=9178264;
RA Papanastasiou P., Bruderer T., Li Y., Bommeli C., Koehler P.;
RT "Primary structure and biochemical properties of a variant-specific
RT surface protein of Giardia.";
RL Mol. Biochem. Parasitol. 86:13-27(1997).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97333006; PubMed=9078242;
RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
RT "The variant-specific surface protein of Giardia, VSP41, is a
RT glycosylated and palmitoylated protein.";
RL Biochem. J. 332:49-56(1997).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -1- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
CC AT THE REDUCING TERMINUS.
CC -1- PTM: PALMITOYLATED.

```

```

CC -----
CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z83743; CAB06038.1; -
DR HSSP: O14763; IDOG.
DR GlycoSuiteDB: P92127; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR005127; Giardia_VSP.
DR Pfam: PF03302; VSP; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF_Like; 1.
DR SMART; SM00261; FU_3.
KW Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
KW Signal.
FT SIGNAL 1 14
FT CHAIN 15 687 POTENTIAL.
FT DOMAIN 15 660 VARIANT-SPECIFIC SURFACE PROTEIN VSP41.
FT TRANSMEM 661 681 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 682 687 POTENTIAL.
FT DOMAIN 687 687 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Alignment Scores:
Pred. No.: 23.3 Length: 687
Score: 55.50 Matches: 12
Percent Similarity: 41.03% Conservatve: 4
Best Local Similarity: 30.77% Mismatches: 8
Query Match: 26.68% Indels: 15
DB: 1 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x VS41_GIALA (1-687)

QY 12 TTTTGCTTGCAGAGATCTTTTCATCTTGCAGG-----
Db 500 TyrCySAlaAlaGlyPheLeuTyrMetGlyGlyCysTyrLysIleAspThrValPro 519
QY 48 -----ACTTCTGGGGCGGAGTAGTAACACTCGGCTC 83
Db 520 GlySerTyrMetCysSerLysSerThrThrAlaGlyValCysAspThrProAsnAla 538

RESULT 12
N12B_MEDSA
ID N12B_MEDSA STANDARD: PRT; 113 AA.
AC Q40339; Q40342;
DT 15-DEC-1998 (Rel. 37; Created)
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DT 15-DEC-1998 (Rel. 37; Last annotation update)
DE Early nodulin 12B precursor (N-12B).
GN ENOD12B.
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
OX NCBI_TaxID=3879;
RN RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nagyszelesi;
RX MEDLINE=93144711; PubMed=7678770;
RA Allison L.A., Kiss G.B., Bauer P., Poliet M., Pierre M.,
RT Savoure A., Kondorosi E., Kondorosi A.;
RT "Identification of two alfalfa early nodulin genes with homology to
RT members of the pea Enod12 gene family.";
RL Plant Mol. Biol. 21:375-380(1993).
RN [2]
RP SEQUENCE OF 15-113 FROM N.A.
RC STRAIN=cv. COERULA W2;

```

```

RT other properties.";
RN [2].
RL Eur. J. Biochem. 230:977-986(1995).
RR
RP SEQUENCE OF 4-150 FROM N.A.
RC TISSUE=Liver;
RA Archibald A.L., Couperwhite S., Jiang Z.H.;
RL "The porcine TTR locus maps to chromosome 6q.";
RT "The porcine TTR locus maps to chromosome 6q.";
CC -1- FUNCTION: Thyroid hormone-binding protein. Probably transports
CC thyroxine from the bloodstream to the brain.
CC -1- SUBUNIT: Homotrimer.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED AND SECRETED BY THE CHOROID
CC PLEXUS, IT IS ALSO PRESENT IN THE LIVER OF BIRDS AND EUTHERIANS.
CC -1- SIMILARITY: BELONGS TO THE TRANSTHYRETIN FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; X82258; CAA57713.1; -.
DR EMBL; U16131; AAA79042.1; -.
DR EMBL; X87846; CAA61120.1; -.
DR HSSP; P02767; IGRE
DR InterPro; IPR000895; Transthyretin.
DR Pfam; PF00576; Transthyretin.1.
DR PRINTS; PR00189; TRANSTHYRETIN.
DR ProDom; PD003457; Transthyretin.1.
DR SMART; SM00095; TR_THY.1.
DR PROSITE; PS00768; TRANSTHYRETIN.1; 1.
DR PROSITE; PS00769; TRANSTHYRETIN.2; 1.
DR Albumin; Transport; Retinol-binding; Vitamin A; Thyroid hormone;
KW Signal.
KW FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 150 TRANSTHYRETIN.
FT BINDING 36 36 THYROID HORMONES (BY SIMILARITY).
FT BINDING 74 74 THYROID HORMONES (BY SIMILARITY).
SQ SEQUENCE 150 AA; 16081 MW; 0427ED5D3094CA07 CRC64;

Alignment Scores:
Pred. No.:
Score: 31.4 Length: 150
Percent Similarity: 54.50 Matches: 12
Best local Similarity: 62.50% Conservative: 3
Query Match: 50.00% Mismatches: 8
Db: 26.20% Indels: 1
Gaps: 1

US-09-513-999C-3792_COPY_51_161 (1-111) x TTHY_PIG (1-150)
QY 9 ATCTTTGCGCTTGAGGATCTTTTCATCTT--GCAGGAGCTGCGGGCCGAGTA 65
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8 leuLeucylsleuAlaIagIleuValAlpheValserGlnAlaIagIleuValAlaIagIleu 27
QY 66 TGTAAACTCT 77
||| |||
Db 28 serIyscysPro 31

RESULT 14
UT2_RAT
AC 062668; STANDARD; PRT; 397 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE urea transporter, kidney.
GN SIC1A2 OR UT2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;

```



```

collagen alpha 3(IV) chain - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
R:Morrisson, K.E.; Germino, G.G.; Keeders, S.T.
J: Biol. Chem. 266, 34-39, 1991
J:Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the
A:Reference number: A39024; MUID:91093146; PMID:1985905
A:Accession: A39024
A:Molecule type: mRNA
A:Residues: 1-471 <MO>
A:Cross-references: EMBL:M63139; NID:9162886; PID:AA62708.1; PID:9162887
R:Butkovski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
J: Biol. Chem. 262, 7874-7877, 1987
A:Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
A:Reference number: S18432; MUID:87222419; PMID:2438283
A:Accession: S20672
A:Molecule type: Protein
A:Residues: 227-228, 'X', 230-244 <BUT>
R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
J: Biol. Chem. 263, 13374-13380, 1988
A:Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen
A:Reference number: S17802; MUID:88330844; PMID:3417661
A:Accession: S17802
A:Molecule type: Protein
A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
J: Biol. Chem. 265, 5466-5469, 1990
A:Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type
A:Reference number: A35167; MUID:90202779; PMID:2318822
A:Accession: A35167
A:Molecule type: Protein
A:Residues: 236-258 <GUN>
R:Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe
J: Biol. Chem. 266, 15318-15324, 1991
A:Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll
A:Reference number: A39419; MUID:91332055; PMID:1869555
A:Accession: C39419
A:Molecule type: Protein
A:Residues: 236-255 <GU2>
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication;
E:1-238/Domain: collagenous (fragment) #status predicted <CO>
E:239-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
E:239-353/Domain: repeat NC1 #status predicted <NC1>
E:354-471/Domain: repeat NC1 #status predicted <NC1>
F:232, 238/Modified site: hydroxyproline (Pro) #status experimental
F:306-312, 417-423/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.: 9.67 length: 471
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
Db: 2 Gaps: 3

US-09-513-999C-3792_COPY_51_161 (1-111) x A39024 (1-471)

OY 3 GGGTGGATCTTTTCCCTTGACGATCTTTTCATCTTT-----GCAGGAGCTTCT 53
|||||
Db 374 G|YTPRILSERLEUTRTP|YSGLYPHESRPH|LEMETPHER|HSR|ALAG|YSERGLU 393
|||||
OY 54 GGGGCCGGA---GTAATGTAACCTCGGCGCTCTGTGCTGCTGCGTAGTG 101
|||||
Db 394 G|YLAAG|G|NALAL|A|L|A|S|ER|P|RO|GLY|S|E|---C|YS|LE|G|L|U|P|H|E 409
|||||

RESULT 3
T16300
hypothetical protein FB86.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16300

```

```

R:WILCOX, L.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F38B6.
A:Reference number: Z18491
A:Accession: T16300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-314 <NID>
A:Cross-references: EMBL:U40060; NID:g1055174; PID:g1055179; PIDN:AAA81145.1; CESP:F38B6.7
A:Gene: CESP:F38B6.7
A:Introns: 44/1; 70/1; 120/1; 179/1; 219/3; 258/3

Alignment Scores:
Pred. No.: 11.5 Length: 314
Score: 60.00 Matches: 11
Percent Similarity: 62.96% Conservative: 6
Best Local Similarity: 40.74% Mismatches: 6
Query Match: 28.85% Indels: 4
DB: 2 Gaps: 1

US-09-513-999C-3792_COPY_51_161 (1-111) x T16300 (1-314)

QY 11 CTTTTCCTTCGACGAGATTCTTTTCATCTT-----TGCAGGAGACTTCGGGCG 58
:::||||| ||| :::::||||| ||| |||:::||||| |||
Db 278 ILETLuLscYsgLmEILepheMetLeuILeThrLeuHISLeuARgAsnPhetRpLly 297
CGGAGTATGTAAACTCCTGG 79

QY 59 CGGAGTATGTAAACTCCTGG 79
||| ::::: |||::: |||
Db 298 ArGLysLeuGLuAsnAsnTrp 304

RESULT 4
T15346
cellulose synthase catalytic subunit-like protein - Arabidopsis thaliana
N:Alternate names: protein F2K13_60
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T15346
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.;
submitted to the Protein Sequence Database, August 2000
A:Reference number: Z25394
A:Accession: T15346
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1145 <SAT>
A:Cross-references: EMBL:AL391141
A:Experimental source: cultivar Columbia; BAC clone F2K13
C:Genetics:
A:Map position: 5
A:Introns: 297/2; 566/3
A:Note: F2K13_60

Alignment Scores:
Pred. No.: 10.7 Length: 1145
Score: 60.00 Matches: 13
Percent Similarity: 57.14% Conservative: 3
Best Local Similarity: 46.43% Mismatches: 11
Query Match: 28.85% Indels: 1
DB: 2 Gaps: 0

US-09-513-999C-3792_COPY_51_161 (1-111) x T15346 (1-1145)

QY 8 GATCTTTGGCTTCGAGGATTCCTTTTCATCTTTTCGACGAGACTTCGGGCGGAGATG 67
||||| ||| ||| ::: ||| ||| ||| ::: ||| ||| |||
Db 145 AspleuLeuProCysGLuCyAspPheLysILeCySarGAspCysPhe-ILeAspAlaVa 164
TAAACTCCTGGGTCCTGTGT 89
||||| ||| ::: |||
Db 164 ILysThrELyGLyLILeCys 171

RESULT 5
T15348
tylpothetical protein B0350.1 - Caenorhabditis elegans

```

C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T15348
 R:Gattung, S.
 Submitted to the EMBL Data Library, February 1996
 A:Description: The sequence of C. elegans cosmid B0350.
 A:Reference number: Z18332
 A:Accession: T15348
 A:Status: preliminary; translated from GB/EMBL/DDJ
 A:Molecule type: DNA
 A:Residues: 1-5170 <GAT>
 A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208877; PIDN:AAA93447.1; CESP:B0350
 C:Genetics:
 A:Gene: CESP:B0350.1
 A:Introns: 48/71; 5039/3; 5116/3

Alignment Scores:
 Pred. No.: 9.85 Length: 5170
 Score: 60.00 Matches: 11
 Percent Similarity: 51.85% Conservative: 3
 Best Local Similarity: 40.74% Mismatches: 13
 Query Match: 29.70% Indels: 0
 DB: 2 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x T15348 (1-5170)
 QY 107 GAGCAGCCACTCAGCAGACAGACAGAGGTTTACATACCGCCAGAGT 48
 Db 972 GUGerProleglunhsglunhsglunhglunhphshtserhshProgluser 991
 QY 47 CCCTGCAGAGATGAAAAAGAA 27
 Db 992 Provalleuserglunysgu 998

RESULT 6
 B97583
 hypothetical protein AGR_C_3398 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: B97583
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappes, C.; Markelz, B.;
 Science 294, 2333-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; PMID:11743194
 A:Accession: B97583
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-382 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK07619.1; PID:g15156967; GSPDB:GN00169
 C:Genetics:
 A:gene: AGR_C_3398
 A:Map position: circular chromosome

Alignment Scores:
 Pred. No.: 15.5 Length: 382
 Score: 59.00 Matches: 15
 Percent Similarity: 43.90% Conservative: 3
 Best Local Similarity: 36.59% Mismatches: 13
 Query Match: 28.37% Indels: 10
 DB: 2 Gaps: 2

US-09-513-999c-3792_COPY_51_161 (1-111) x B97583 (1-382)
 QY 6 TGGAGCTTTGGCTTCGAGAGTCTTTTCATCTTT-----GCA 44
 Db 116 TTPARGPhe-----GlyLeuphepellepheProphleuLeuMetLeuAlaGly 132
 QY 45 GGGACTTCTGGGGCCGAGATATGTAACCTCCTGGCTCTGTGTGTCCTGAGTGGCTG 104
 Db 133 GlyAlaLeuAlaAlaGlyIleAlaLeuSerProtrPleuPheSerLeuProleuTripleu 152
 QY 105 CTC 107

Db 153 Leu 153

RESULT 7
 AB2804
 hypothetical protein Atui852 [imported] - Agrobacterium tumefaciens (strain C58, Dupo
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AB2804
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo
 erage, G.; Gillet, W.; Grant, C.; Guentherer, D.; Kutayarin, T.; Levy, R.; Li, M.; MccI
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AB2804
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-382 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAL42848.1; PID:g17740297; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atui852
 A:Map position: circular chromosome

Alignment Scores:
 Pred. No.: 15.5 Length: 382
 Score: 59.00 Matches: 15
 Percent Similarity: 43.90% Conservative: 3
 Best Local Similarity: 36.59% Mismatches: 13
 Query Match: 28.37% Indels: 10
 DB: 2 Gaps: 2

US-09-513-999c-3792_COPY_51_161 (1-111) x AB2804 (1-382)

QY 6 TGGAGCTTTGGCTTCGAGAGTCTTTTCATCTTT-----GCA 44
 Db 116 TTPARGPhe-----GlyLeuphepellepheProphleuLeuMetLeuAlaGly 132
 QY 45 GGGACTTCTGGGGCCGAGATATGTAACCTCCTGGCTCTGTGTGTCCTGAGTGGCTG 104
 Db 133 GlyAlaLeuAlaAlaGlyIleAlaLeuSerProtrPleuPheSerLeuProleuTripleu 152
 QY 105 CTC 107
 Db 153 Leu 153

RESULT 8
 S25714
 son-of-sevenless-2 protein - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
 C:Accession: S25714
 R:Bowtell, D.; Fu, P.; Simon, M.; Senior, P.
 Proc. Natl. Acad. Sci. U.S.A. 89, 6511-6515, 1992
 A:Title: Identification of murine homologues of the Drosophila Son of sevenless gene:
 A:Reference number: S25714; MUID:9235328; PMID:1631150
 A:Accession: S25714
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1297 <BOW>
 A:Cross-references: EMBL:Z11664; NID:g54136; PIDN:CAA77732.1; PID:g54137
 F:405-507/Domain: plectestrin repeat homology <PLK>
 F:739-882/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Alignment Scores:
 Pred. No.: 14.5 Length: 1297
 Score: 59.00 Matches: 13
 Percent Similarity: 48.57% Conservative: 4
 Best Local Similarity: 37.14% Mismatches: 12

Query Match:	29.21%	Indels:	6
DB:	2	Gaps:	1
US-09-513-999C-3792_COPY_51_161 (1-111) x S25714 (1-1297)			
OY	106	AGCAGCCATCAGGCACACAGACCCAGAGCTTTTACATCTCGGCCGAGAGTC	47
DB	1244	SerSerHisSer-----SerLeuAlaHisLeuProAlaProProVal	1257
OY	46	CCTGCAGAGATGAAAAAGATCTCTGCAGAGCAAGATCCACCCA	2
DB	1258	ProProArgGlnAsnSerSerProLeuLeuProLysLeuProPro	1272
RESULT 9			
T04782			
hypoetical protein F10M10.130 - Arabidopsis thaliana			
C:Species: Arabidopsis thaliana (mouse-ear cress)			
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999			
C:Accession: T04782			
R:By:Van, M.; De Haan, M.; Maarse, A.C.; Griwell, L.A.; Hohelsel, J.; Nemes, H.W.; Mayer, submitted to the Protein Sequence Database, February 1999			
A:Reference number: 215384			
A:Accession: T04782			
A:Molecule type: DNA			
A:Residues: 1-197 <BEV>			
A:Cross-references: EMBL:AL035521			
A:Experimental source: cultivar Columbia; BAC clone F10M10			
C:Genetics:			
A:Map position: 4			
A:Introns: 49/1; 76/3; 121/3; 156/3			
A:Note: F10M10.130			
Alignment Scores:			
Pred. No.:	21.9	Length:	197
Score:	58.00	Matches:	14
Percent Similarity:	60.87%	Conservative:	0
Best Local Similarity:	60.87%	Mismatches:	8
Query Match:	28.71%	Indels:	1
DB:	2	Gaps:	0
US-09-513-999C-3792_COPY_51_161 (1-111) x T04782 (1-197)			
OY	72	TTTTNATCTCCGCCGCCAGAGCCCGCAAGATGAAAGAAATCCTGCAGGCAAA	13
DB	151	PhETyTtLeuATgLySgLyLysArg-CysAsnAspGluLysGluAspGluLysCysAs	170
OY	12	AGATCCA	6
b	170	naspro	172
RESULT 10			
AA4356			
calhepsin E (EC 3.4.23.34) precursor - guinea pig			
N:Alternate names: erythrocyte membrane aspartic proteinase; slow-moving proteinase			
C:Species: Cavia porcellus (guinea pig)			
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999			
C:Accession: AA4356			
R:Kageyama, T.; Ichinose, M.; Tsukada, S.; Miki, K.; Kurokawa, K.; Koivai, O.; Tanji, M			
J. Biol. Chem. 267, 16450-16459, 1992			
A:Title: Gastric procalhepsin E and progastricsin from guinea pig. Purification, molecu			
A:Reference number: AA4356; MUID:92355614; PMID:1644829			
A:Accession: AA4356			
A:Molecule type: mRNA			
A:Residues: 1-391 <KAG>			
A:Cross-references: GB:M86653; NID:g191294; PIDN:AAA37052.1; PID:g191295			
A:Note: sequence extracted from NCBI backdome (NCBIN:110763, NCBIp:110769)			
C:Superfamily: pepsin			
C:Keywords: aspartic proteinase; hydrolase; membrane protein			
Alignment Scores:			
Pred. No.:	21.1	Length:	391
Score:	58.00	Matches:	12
Percent Similarity:	63.33%	Conservative:	7

```

Best Local Similarity: 40.00%      Mismatches: 11
Query Match: 27.88%              Indels: 0
DB: 2                             Gaps: 0

US-09-513-999C-3792_COPY_51_161 (1-111) x A43356 (1-391)

QY 1 ATGGGTGATCTTTTGGCTTCGACGATTCCTTTTCATCTTTGACGGGACCTTCGGGCCG 60
    ::||| |||||:::||||| ||||| ||||| |||||:::|||||
DB 127 VALGlyAsnSerPheSerIleGlnTyrGlyThrGlySerLeuThrGlyIleGlyAla 146
    ::||| ::|||:::||||| |||

QY 61 GAGTATGTAAACCTCTGCGTCTCTGTGTG 90
    ::||| ::|||:::||||| |||
DB 147 AspGlnValSerValGlnGlyLeuThrVal 156
    ::||| ::|||:::||||| |||

RESULT 11
T12191
MADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Simulium vittatum mitochondri
C:Species: mitochondion Simulium vittatum
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
C:Accession: T12191
R:Zhu, X.; Pruess, K.P.; Powers, T.O.
submitted to the EMBL Data Library, September 1997
A:Description: Mitochondrial DNA polymorphism in a black fly, Simulium vittatum (Dipt
A:Reference number: Z17447
A:Accession: T12191
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-240 <ZHU>
A:Cross-references: EMBL:AF023535; NID:G2547181; PID:G2547182; PIDN:AAB81263.1
A:Experimental source: strain cytospecies IIL-1 Nebraska
C:Genetics:
A:genome: mitochondrion
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Alignment Scores:
Pred. No.: 29.5      Length: 240
Score: 57.00      Matches: 15
Percent Similarity: 52.94%      Conservative: 3
Best Local Similarity: 44.12%      Mismatch: 16
Query Match: 27.40%      Indels: 0
DB: 2              Gaps: 0

US-09-513-999C-3792_COPY_51_161 (1-111) x T12191 (1-240)

QY 1 ATGGGTGATCTTTTGGCTTCGACGATTCCTTTTCATCTTTGACGGGACCTTCGGGCCG 60
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 196 IlegIglYSerPheLeuSerTrPleuMetPheProSerValSerMetIleCysLeuPro 215
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GAGTATGTAAACCTCTGCGTCTCTGTGTGCTTCGACGATGCGC 102
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||
DB 216 MetTyrTrTylAsnLeuThrLeuPheValCysLeuGlyGly 229
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||

RESULT 12
T04523
hypoetical protein F16A16.130 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
R:Bevan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schon, O.; Hohensei, J.; M
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15376
A:Accession: T04523
A:Molecule type: DNA
A:Residues: 1-880 <BEV>
A:Cross-references: EMBL:AL035353
A:Experimental source: cultivar Columbia; BAC clone F16A16
C:Genetics:
A:Map position: 4
A:introns: 660/1
A:Note: F16A16.130

Alignment Scores:

```


Pred. No.: 27.4 Length: 880
Score: 57.00 Matches: 10
Percent Similarity: 67.86% Conservative: 9
Best Local Similarity: 35.71% Mismatches: 9
Query Match: 28.22% Indels: 0
DB: 2 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x T04523 (1-880)

QY 90 CACAGAGACCCAGGAGTTTACATACCTGCGCCAGACCCGCGCAAGATGAAAA 31

Db 189 HisserlunAlaYsArgIleThrValleuArgProSerIylsAlaYciunhrluYs 208

QY 30 AGAATCTGCAAGGCAAAAGATCC 7

Db 209 TyrValValGlnGlyArgArgAsn 216

RESULT 13

Q08B8 BPLP1 protein - human herpesvirus 4 (strain B95-8)

C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 16-Jul-1999

C:Accession: G93065; A03747; S32993

R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus

A:Reference number: A93065; MUID:85035713; PMID:6092825

A:Accession: G93065

A:Molecule type: DNA

A:Residues: 1-3149 <BAN>

A:Cross-references: EMBL:V01555; NID:959074; PIDN:CA24839.1; PID:91334853

R:Bankier, A.T.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H

Nature 310, 207-211, 1984

A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A:Reference number: A03794; MUID:84270667; PMID:6087149

C:Contents: annotation: protein coding region

C:Superfamily: human herpesvirus 4 BPLP1 protein

Alignment Scores: 25.5 Length: 3149

Pred. No.: 57.00 Matches: 11

Score: 41.18% Conservative: 3

Percent Similarity: 32.35% Mismatches: 20

Best Local Similarity: 28.22% Indels: 0

Query Match: 1 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x Q08B8 (1-3149)

QY 103 AGCCACCTCAGGACACACAGACCCGAGGTTTACATCTCGGCCCCAGATCCCT 44

Db 449 ThrHisValProProHisArgProSerAlaAlaArgLeuProProValIlePro 468

QY 43 GCAAGATGAAAGAAATCGCGCAAGGCAAAAGATCCACCA 2

Db 469 IleProHisGlnSerProProAlaSerProThrProHisPro 482

RESULT 14

A56731 Chromatin assembly factor I p150 chain - human

C:Species: Homo sapiens (man)

C:Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 21-Jul-2000

C:Accession: A56731

R:Kaufman, P.D.; Kobayashi, R.; Kessler, N.; Stillman, R.

Cell 81, 1105-1114, 1995

A:Title: The p150 and p60 subunits of chromatin assembly factor I: a molecular link betw

A:Reference number: A56731; MUID:95333966; PMID:7600578

A:Accession: A56731

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-938 <KAW>

A:Cross-references: GB:U20979; NID:9882257; PIDN:AAA76736.1; PID:9882258

Alignment Scores:

Pred. No.: 31.8 Length: 938
Score: 56.50 Matches: 16
Percent Similarity: 44.64% Conservative: 9
Best Local Similarity: 28.57% Mismatches: 10
Query Match: 27.97% Indels: 21
DB: 2 Gaps: 4

US-09-513-999c-3792_COPY_51_161 (1-111) x A56731 (1-938)

QY 106 AGCAGCCTCAGGACACACAGACCCGAGGATTT-----TTTCATATCTCGGCCCA 53

Db 749 SerHisHisThrIleProSerProAlaGlnProSerThrThrTyrLeuHisThrProThrPro 768

QY 52 GAA-----GTCCCTGCAAG-----ATGAAAAAGATCCT 23

Db 769 SerGluAspAlaAlaIleProSerIylsSerArgLeuYsArgLeuIleSerGluAsnSer 788

QY 22 GCAAGGCAAAAG-----ATCCACCA 2

Db 789 ValTyrGluYsArgProAspPheArgMetCysTyrTyrValHisPro 804

RESULT 15

CGH03B collagen alpha 3(IV) chain precursor, long splice form - human

N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice for

C:Species: Homo sapiens (man)

C:Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999

C:Accession: A54763; A43928; A44043; A45971; A39786

R:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reekers, S.T.

J. Biol. Chem. 269, 23013-23017, 1994

A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpress

A:Reference number: A54763; MUID:94364994; PMID:8083201

A:Accession: A54763

A:Molecule type: mRNA

A:Residues: 1-1670 <MAR>

A:Cross-references: GB:X80031; NID:9577563; PID:9577564

A:Experimental source: kidney

R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.

J. Clin. Invest. 89, 592-601, 1992

A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the

A:Reference number: A43928; MUID:92147878; PMID:1737849

A:Accession: A43928

A:Molecule type: mRNA

A:Residues: 1331-1524, 'T', 1526-1670 <TUR>

A:Cross-references: GB:M81379

A:Experimental source: kidney

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 267, 19780-19784, 1992

A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpas

tion.

A:Reference number: A44043; MUID:93015826; PMID:1400291

A:Accession: A44043

A:Molecule type: DNA; mRNA

A:Residues: 1386-1670 <OUT>

A:Cross-references: GB:M92993; NID:9177895; PIDN:AAA21610.1; PID:9177896

A:Note: Sequence extracted from NCBI backbone (NCBIP:115597)

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 269, 17358, 1994

A:Reference number: A44738; MUID:94274734; PMID:8006044

A:Contents: annotation: error: correction to intronic sequence in A44043

A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.

J. Biol. Chem. 268, 12090-12094, 1993

A:Accession: A45971; MUID:93280184; PMID:8505332

A:Accession: A45971

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1427-1444 <BER>

A:Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly id

R:Morrisson, K.E.; Maruyama, M.; Yang-Feng, T.L.; Reekers, S.T.

Am. J. Hum. Genet. 49, 545-554, 1991

A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain

A:Reference number: A39786; MUID:91335370; PMID:1882840

A:Accession: A39786
 A:Molecule type: mRNA
 A:Residues: 1453-1593, 'A', 1595-1670 <MOR>
 A:Cross-references: GB:S55790; NID:9234418; PIDN:AB19637.1; PID:9234419
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (ed and subsequently O-glycosylated).
 C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope C:genetics:
 A:Gene: GDB:COL4A3
 A:Cross-references: GDB:128351; OMIM:120070
 A:Map position: 2q36-q37
 A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete
 A>Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3(many trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a er associations in the interrupted helical domain (with disulfide and desmosine cross-l
 C:Function:
 A:Description: minor structural component of extracellular basement membrane in kidney g
 C:Superfamily: collagen alpha 1(IV) chain
 Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
 F:1-28/Domain: signal sequence #status predicted <Sig>
 F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
 F:29-42/Domain: amino-terminal nonhelical, NHI <NHI>
 F:43-1438/Region: interrupted helical
 F:791-793/Region: cell attachment (R-G-D) motif
 F:996-998/Region: cell attachment (R-G-D) motif
 F:1154-1156/Region: cell attachment (R-G-D) motif
 F:1306-1308/Region: cell attachment (R-G-D) motif
 F:1345-1347/Region: cell attachment (R-G-D) motif
 F:1432-1434/Region: cell attachment (R-G-D) motif
 F:1439-1670/Domain: carboxyl-terminal nonhelical, NCI <NCI>
 F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F:31,33,39,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status predi
 F:253/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted
 F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Alignment Scores:
 Pred. No.: 30.8 Length: 1670
 Score: 56.50 Matches: 16
 Percent Similarity: 54.05% Conservative: 4
 Best Local Similarity: 43.24% Mismatches: 12
 Query Match: 27.16% Indels: 5
 DB: 1 Gaps: 3

US-09-513-999C-3792_COPY_51_161 (1-111) x CGH3B (1-1670)
 QY 3 GGGTGAATCTTTGCGAGATTCCTTTTCATCTT-----GCAGGGAATCTT 53
 Db 1573 GLYTRPLESERLEUTRPLYSGLYPHERPHERPHERSERALAGYSEGLU 1592
 QY 54 GGGGCCGGA---GATGTAAACTCCGCGCTCTGCTGCTGCTGCTGCTGCTG 101
 Db 1593 GLYTRGLYSINALALALALASERPROGLYSER---CYSLEUGLUGLUPHE 1608

Search completed: April 30, 2003, 14:36:50
 Job time : 17.5 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 30, 2003, 14:34:50 ; Search time 32.5 Seconds

(without alignments)
910.204 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161

Perfect score: 208 1 atgggttgagctctttgctt.....gctcgtggtgctctctact 111

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ .n2p.model -DEV=xlp
-O=/cgn2_1/uspro.spool/us0951399/rtunat_30042003_143400_25586/ftp_query.fasta_1.263
-DB-A-Geneseq_101002 -OPMT=fastan -SUFFIX=n2p.rag -MIMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsun62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pcio -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09513999_@cgn2_1.114_@rtunat_30042003_143400_25586 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEIOUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A.Geneseq_101002: *
1: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1980.DAT: *
2: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1981.DAT: *
3: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1982.DAT: *
4: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1983.DAT: *
5: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1984.DAT: *
6: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1985.DAT: *
7: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1986.DAT: *
8: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1987.DAT: *
9: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1988.DAT: *
10: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1989.DAT: *
11: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1990.DAT: *
12: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1991.DAT: *
13: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1992.DAT: *
14: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1993.DAT: *
15: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1994.DAT: *
16: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1995.DAT: *
17: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1996.DAT: *
18: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1997.DAT: *
19: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1998.DAT: *
20: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1999.DAT: *
21: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2001.DAT: *
23: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	92.8	131	21	AAAG03788
2	91	45.0	280	22	ABG15067
3	85	40.9	57	23	ABP09726
4	80	39.6	352	22	ABG27640
5	78	37.5	93	22	AAW24321
6	67	33.2	864	22	ABG07580
7	67	33.2	864	22	AAU32809
8	63.5	31.4	386	22	AAU19510
9	63.5	31.4	640	20	AAZ99918
10	63.5	31.4	648	20	AAZ99917
11	63.5	31.4	648	22	ABG95518
12	63.5	31.4	1176	22	ABG22576
13	61	29.3	52	22	AAW24448
14	60.5	29.1	77	22	AAW48057
15	60.5	29.1	218	16	AAW79164
16	60.5	29.1	218	20	AAW44172
17	60.5	29.1	218	21	AAW56784
18	60.5	29.1	218	22	AAW09484
19	60.5	29.1	268	20	AAW31993
20	60.5	29.1	268	21	AAW97555
21	60.5	29.1	471	16	AAW79163
22	60.5	29.1	471	20	AAW44171
23	60.5	29.1	471	21	AAW56783
24	60.5	29.1	471	22	AAW09483
25	60	28.8	1145	23	ABW93513
26	59	29.2	854	22	ABG28918
27	59	29.2	1297	14	AAW47048
28	59	29.2	1297	16	AAW84639
29	58.5	29.0	126	22	AAW92343
30	58.5	29.0	583	22	ABW71010
31	58.5	29.0	668	20	AAW33029
32	58	27.9	299	21	AAW53137
33	58	27.9	887	22	ABG23684
34	58	28.7	1099	22	ABW68903
35	58	27.9	1113	22	ABG06702
36	57	27.4	61	22	AAW01610
37	57	28.2	505	22	AAW80430
38	57	28.2	509	22	AAW62173
39	57	28.2	509	22	AAW80382
40	57	28.2	509	23	ABG63574
41	57	28.2	509	23	ABG65272
42	57	28.2	554	22	AAW80407
43	57	28.2	554	23	ABG65271
44	57	28.2	594	22	AAW93619
45	57	28.2	783	20	AAW60344

ALIGNMENTS

RESULT 1
ID AAG03788 standard: Protein; 131 AA.

XX AAG03788:

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 7869.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping.

XX Homo sapiens.

OS EP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 DR N-PSDB: AAC03794.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13; SEQ ID 7869; 71pp + CD-ROM; English.
 XX
 CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX
 SQ Sequence 131 AA:
 Alignment Scores:
 Pred. No.: 1.58e-16 Length: 131
 Score: 193.00 Matches: 37
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 92.79% Indels: 0
 DB: 21 Gaps: 0
 US-09-513-999c-3792_COPY_51_161 (1-111) x AAG03788 (1-131)
 QY 1 ATGGTGATCTTTTGCCTTCAGAGATCTTTTCATCTTTCAGAGGACTTCGGGCGG 60
 DB 1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
 QY 61 GAGTATGTAACATCCTGGGTCTGTGCTGTGCTGAGTGGCTGCTCTACT 111
 DB 21 GlutryValLysLeuLeuGlyLeuGlyValCysLeuSerGlyCysSerThr 37
 RESULT 2
 ABG15067
 ID ABG15067 standard; Protein: 280 AA.
 XX
 AC ABG15067;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #15058.
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001MO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639363/73.
 DR N-PSDB: AAS79234.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 45426; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 280 AA:
 Alignment Scores:
 Pred. No.: 0.00247 Length: 280
 Score: 91.00 Matches: 17
 Percent Similarity: 63.64% Conservative: 4
 Best Local Similarity: 51.52% Mismatches: 12
 Query Match: 45.05% Indels: 0
 DB: 22 Gaps: 0
 US-09-513-999c-3792_COPY_51_161 (1-111) x ABG15067 (1-280)
 QY 100 CACTCAGGACACACAGACAGACCCAGATTTACATACCTCCGCCACAACTCCCTGCA 41
 DB 23 HisSerAspThrTyrArgTyrLeuGlyAlaLeuAspThrProAlaProValPheProThr 42
 QY 40 AAGATGAAAAGATCCTCAGAGGCAAGGCAAGATCCACCA 2
 DB 43 LysValSerAlaThrProAlaLysGlnLysValAlaGPro 55
 RESULT 3
 ABP09726
 ID ABP09726 standard; Protein: 57 AA.
 XX
 AC ABP09726;
 XX
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX protein sequence SEQ ID NO:19434.
 XX
 KW Human: open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;

KV		immune deficiency; immune disorder; infectious disease;
KW		autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KX		myasthenia gravis.
XX		
OS	Homo sapiens.	
PN	WO200192523-A2.	
PP	06-DEC-2001.	
PB	29-MAY-2001; 2001WO-US10836.	
PR	30-MAY-2000; 2000US-206132P.	
PR	29-AUG-2000; 2000US-228716P.	
XX		
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Shinkets RA, Leach MD;	
DR	WPI: 2002-106308/14.	
JR	N-PSTDB: ABN25478.	
XX		
PT	Novel human polypeptides and polynucleotides useful for diagnosing,	
PT	preventing and treating cardiovascular disease, neurodegenerative,	
PT	hyperproliferative disorders and autoimmune disorders -	
XX		
PS	Disclosure; SEQ ID 19434; 1037bp; English.	
CC		
CC	The present invention describes substantially purified human proteins	
CC	(referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1	
CC	in the specification), ABN15762 to ABN27252 encode the human ORFX	
CC	proteins given in ABP00010 to ABP11500. ORFX proteins are useful for	
CC	treating or preventing a pathology associated with an ORFX-associated	
CC	disorder in humans, and in the manufacture of a medicament for treating	
CC	syndrome associated with ORFX-associated disorder. ORFX polynucleotide	
CC	sequences can be used in gene therapy. ORFX sequences can be used in the	
CC	treatment of cancer, hyperproliferative disorders, cirrhosis of liver,	
CC	psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,	
CC	osteoarthritis, neurodegenerative disorders, disorders related to organ	
CC	transplantation, cardiovascular diseases, diabetes mellitus, systemic	
CC	lupus erythematosus, hypertension, hypothyroidism, cholesterol ester	
CC	storage disease, various immune deficiencies and disorders, infectious	
CC	diseases, autoimmune disorders such as multiple sclerosis, Rheumatoid	
CC	arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host	
CC	disease and autoimmune inflammatory eye disease. ORFX proteins are also	
CC	useful for treating burns, incisions, ulcers, for treating osteoporosis,	
CC	bone degenerative disorders, or periodontal disease, and for gut	
CC	protection or regeneration and treatment of lung or liver fibrosis,	
CC	repulsion injury in various tissues and conditions resulting from	
CC	systemic cytokine damage.	
CC	N.B. The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 57 AA;	
Alignment Scores:		
Pred. No:	0.013	Length: 57
Score:	85.00	Matches: 14
Percent Similarity:	94.12%	Conservative: 2
Best Local Similarity:	82.35%	Mismatches: 1
Query Match:	40.87%	Indels: 0
DB:	23	Gaps: 0
US-09-513-999C-3792_COPY_51_161 (1-111) x ABP09726 (1-57)		
OY	58 CCGAGATGTTAAACTCTGGGTCTCTGTCTGTCGCTGAGTGCCTCTCT 108	
Db	20 ProglutinytreunysleuenglylaucyslleucysleueserSerCyser 36	
RESULT 4		
ABG27640		
ID	ABG27640 standard; Protein: 352 AA.	

[illegible]

QY 46 CCTGCAAGATGAAAAAGATCTCGACAGGCAAAAG 11
 DB 210 ProthrAsnValAlaAlaIthrGlnAlaArgArgGlu 221
 RESULT 5
 ID AAM24321 standard; Protein; 93 AA.
 AC AAM24321;
 DT 12-OCT-2001 (first entry)
 DE Human EST encoded protein SEQ ID NO: 1846.
 XX
 XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KM diagnostics; forensic test; gene mapping; genetic disorder;
 KM biodiversity; gene therapy; nutrition.
 XX
 XX Homo sapiens.
 PN WO200154477-A2.
 PD 02-AUG-2001.
 XX
 XX 25-JAN-2001; 2001MO-US02687.
 PF 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 XX
 DR WPI: 2001-476164/51.
 DR N-PSDB; AAB98980.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 20; Page 1198; 1275pp; English.
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 XX
 XX SQ Sequence 93 AA;
 Alignment Scores:
 Pred. No.: 0.108 Length: 93
 Score: 78.00 Matches: 17
 Percent Similarity: 71.43% Conservative: 3
 Best Local Similarity: 60.71% Mismatches: 8
 Query Match: 37.50% Indels: 0
 Gaps: 0
 US-09-513-999c-3792_COPY_51_161 (1-111) x AAM24321 (1-93)
 QY 16 GCCTTGACAGATCTTTTCAATCTTGCAGGACTCTGGGCGCGAGATGTAATAATC 75
 DB 62 AlaIeuProGluMetProSerProPheIeuClyIleIeuArgIeuGluIuTyValIyLsu 81
 QY 76 CTGGGTCTCTGTGTGTGCTGAGT 99
 DB 82 IeuGlyIeuCysMetCysIeuSer 89

RESULT 6
 ID ABG07580 standard; Protein; 864 AA.
 AC ABG07580;
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #7571.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 XX Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001MO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 DR N-PSDB; AAS71767.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 37939; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG03377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC atftp.wipo.int/pub/published_pcl_sequences.
 XX
 XX SQ Sequence 864 AA;
 Alignment Scores:
 Pred. No.: 3.37 Length: 864
 Score: 67.00 Matches: 16
 Percent Similarity: 58.06% Conservative: 2
 Best Local Similarity: 51.61% Mismatches: 13
 Query Match: 33.17% Indels: 0
 Gaps: 0
 US-09-513-999c-3792_COPY_51_161 (1-111) x ABG07580 (1-864)

OY 98 CTCAGGCACACAGAGCCAGAGTTTACATCTCCGCCCCAGAGTCCCTGCACAA 39
 ||||| :||| |||||
 Db 696 LeuArgGlyProProGlnLeuGlnSerGlnLeuValSerGlyProArgSerProLysSer 715
 OY 38 GATGAAAAAGAACTCTGCAGGCAAAAGATCCA 6
 ||| ||||| ||| |||||
 Db 716 AspSerGlyGlnSerCysLeuAlaTrpAspPro 726

RESULT 7
 AAU32809
 ID AAU32809 standard; Protein: 864 AA.
 AC AAU32809;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #3300.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 XX Claim 20; Page 668; 765pp; English.

CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukemias. AAU29510-AAU3304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX

SO Sequence 864 AA:

Alignment Scores:
 Pred. No.: 3.37 Length: 864
 Score: 67.00 Matches: 16
 Percent Similarity: 58.06% Conservative: 2
 Best Local Similarity: 51.61% Mismatches: 13
 Query Match: 33.17% Indels: 0
 DB: 22 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x AAU32809 (1-864)

OY 98 CTCAGGCACACAGAGCCAGAGTTTACATACTCCGCCCCAGAGTCCCTGCACAA 39
 ||||| :||| |||||
 Db 696 LeuArgGlyProProGlnLeuGlnSerGlnLeuValSerGlyProArgSerProLysSer 715
 OY 38 GATGAAAAAGAACTCTGCAGGCAAAAGATCCA 6
 ||| ||||| ||| |||||
 Db 716 AspSerGlyGlnSerCysLeuAlaTrpAspPro 726

RESULT 8
 AAU19510
 ID AAU19510 standard; Protein: 386 AA.
 AC AAU19510;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Human diagnostic and therapeutic polypeptide (DITHP) #96.
 XX
 KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
 KW respiratory disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200162927-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 21-FEB-2001; 2001WO-US06059.
 XX
 PR 24-FEB-2000; 2000US-0184693.
 PR 24-FEB-2000; 2000US-0184697.
 PR 24-FEB-2000; 2000US-0184698.
 PR 24-FEB-2000; 2000US-0184768.
 PR 24-FEB-2000; 2000US-0184769.
 PR 24-FEB-2000; 2000US-0184770.
 PR 24-FEB-2000; 2000US-0184771.
 PR 24-FEB-2000; 2000US-0184772.
 PR 24-FEB-2000; 2000US-0184773.
 PR 24-FEB-2000; 2000US-0184774.
 PR 24-FEB-2000; 2000US-0184776.
 PR 24-FEB-2000; 2000US-0184777.
 PR 24-FEB-2000; 2000US-0184797.
 PR 24-FEB-2000; 2000US-0184813.
 PR 24-FEB-2000; 2000US-0184837.
 PR 24-FEB-2000; 2000US-0184841.
 PR 24-FEB-2000; 2000US-0185213.
 PR 24-FEB-2000; 2000US-0185216.
 PR 12-MAY-2000; 2000US-0203785.
 PR 15-MAY-2000; 2000US-0204226.
 PR 16-MAY-2000; 2000US-0204525.
 PR 16-MAY-2000; 2000US-0204821.
 PR 16-MAY-2000; 2000US-0204908.
 PR 16-MAY-2000; 2000US-0205232.
 PR 17-MAY-2000; 2000US-0204815.
 PR 17-MAY-2000; 2000US-0204863.
 PR 17-MAY-2000; 2000US-0205221.
 PR 17-MAY-2000; 2000US-0205285.
 PR 17-MAY-2000; 2000US-0205286.
 PR 17-MAY-2000; 2000US-0205287.
 PR 17-MAY-2000; 2000US-0205323.
 PR 17-MAY-2000; 2000US-0205324.
 XX

(INCY-) INCYTE GENOMICS INC.
 PA Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 XX Chen A, D'Sa SA, Amesley S, Dahl CR, Dam TC, Daniels SE;
 XX Dutour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo PD, Stockdreher TK, Daffo A;
 PI Wright RJ, Yap PE, Yu JT, Bradley DL, Bratcher SK, Chen W;
 PI Cohen RJ, Hodgson DM, Lincoln SE, Jackson S;

XX	WP1: 2001-502867/55.
DR	N-PSDB: AAS31081.
XX	
XX	Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT	enzymes, hormones and receptors, useful in diagnostics and therapeutics
PT	-
XX	
PS	Claim 27: Page 452-453; 522pp; English.
CC	
CC	The invention relates to polynucleotides (I) encoding diagnostic and
CC	therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC	and proteins involved in growth and development and receptors. (I) and
CC	(II) may be used in the prevention, diagnosis and treatment of diseases
CC	associated with inappropriate DITHP expression. For example, (I) and
CC	(II) may be used to treat disorders associated with decreased polypeptide
CC	expression by rectifying mutations or deletions in a patient's genome,
CC	that affect the activity of the DITHPs, by expressing inactive proteins
CC	or supplementing the patient's own production of them. (I) and (II)
CC	may be used to treat diseases, for example, cell proliferative disorder,
CC	Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC	leukemia, autoimmune disorders, and respiratory disorders. Additionally,
CC	(I) may be used to produce the DITHPs, by inserting the nucleic acids
CC	into a host cell and culturing the cell to express the protein. (I) and
CC	its complementary sequences may also be used as DNA probes in diagnostic
CC	assays to detect and quantitate the presence of similar nucleic acids in
CC	samples, and therefore which patients may be in need of restorative
CC	therapy. (II) may also be used as antigens in the production of
CC	antibodies against DITHPs and in assays to identify modulators of DITHP
CC	expression and activity. The anti-DITHP antibodies and antagonists may
CC	also be used to down regulate expression and activity. The anti-DITHP
CC	antibodies may also be used as diagnostic agents for detecting the
CC	presence of DITHPs in samples (e.g. by enzyme linked immunosorbant
CC	assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and
CC	therapeutic (DITHP) polypeptides of the invention.
CC	
XX	
SO	Sequence 386 AA:
Alignment Scores:	
Pred. No.:	8.98 Length: 386
Score:	63.50 Matches: 16
Percent Similarity:	66.67% Conserves: 8
Best Local Similarity:	44.44% Mismatches: 7
Query Match:	31.44% Indels: 5
DB:	22 Gaps: 3
US-09-513-999C-3792-COPY_51.161 (1-111) x AAU19510 (1-386)	
Y	110 GTACAGCAGCCACTCAGGCACACAGACAGCCAGG-----AGTTTACATACCGCG 57
	: : : : : : : : : : : : :
Db	314 IIEGAGLURP---GInGUthRGInGUProGUllleuSePeRHeThYTTThrGly 332
	: : : : : : : : : : : : :
OY	56 CCCGAAGTCCCTGCAGAAAGATGAAAAAGATTCCTCAGAGGCAAAAGAT 9
	: : : : : : : : : : : : :
Db	333 AsprfGser-----LysAspGInGUInGUcYsLeuGInGUInGUaSp 346
	: : : : : : : : : : : : :
RESULT 9	
ID	AAI29918 standard; Protein; 640 AA.
AC	AAI29918:
XX	
XX	AAI29918:
XX	
DT	18-NOV-1999 (first entry)
XX	
DE	Human CHD1 protein encoded by transcript cDNA2.
XX	
XX	Human: coronary heart disease susceptibility gene; CHD1; mutation;
KM	chromosome 11; diagnosis; screening; PCR primer; metabolic disorder;
KM	detection; hypoparathyroidism; familial combined hyperlipidaemia;
KM	insulin resistant syndrome X; multiple metabolic disorder; obesity;
KM	diabetes; dyslipidaemic hypertension.
XX	
XX	Homo sapiens.
XX	

PX		PN	MW9945112-AZ.	
XX		XX		
PD		PD	10-SEP-1999.	
PF		PF	04-MAR-1999;	99MO-USO4682.
PR		PR	04-MAR-1998;	98US-0034941.
XX		XX	06-APR-1998;	98US-0080934.
PA	(MTR1-)	GENETICS INC.		
PI	Ballingner DG,	Ding W,	Wagner S,	Hess MA:
DR	N-PSDB:	AAM26833.		
PT	New isolated coronary heart disease susceptibility gene, used to develop products for diagnosis and treatment of coronary heart disease and metabolic disorders -			
PS	Claim 10; Page 124-125; 297pp;	English.		
CC	The present invention describes the human chromosome 11-linked coronary heart disease susceptibility gene (CHD1). Mutations in the CHD1 locus in the germline are indicative of a predisposition to coronary heart disease or to metabolic disorders related to lipid metabolism.			
CC	Products from the present invention can be used in the diagnosis of predisposition to coronary heart disease and to metabolic disorders, including hypothyroidism, familial combined hyperlipidaemia, insulin resistant syndrome X or multiple metabolic disorder, obesity, diabetes and dyslipidemic hypertension. CHD1 proteins can be used for treating coronary heart disease and metabolic disorders. The products can also be used for detection and drug screening. AA226832 to AA226841 and AA227027 to AA227029 represent human CHD1 nucleotide sequences.			
CC	AA229917 to AAY29926 represent human CHD1 proteins and protein sequences used in the exemplification of the present invention. AA226842 to CC			
CC	used in the exemplification of the present invention. AA226862 represent primers used in the identification of human CHD1; AA226863 to AA227014 represent PCR primers used in the screening of mutations in human CHD1; AA227015 to AA227026 represent oligonucleotides used in the exemplification of the present invention.			
SQ	Sequence	640 AA;		
Alignment Scores:				
Pred. No.:	9.32	Length:	640	
Score:	63.50	Matches:	16	
Percent Similarity:	66.67%	Conservative:	8	
Best local Similarity:	44.44%	Mismatches:	7	
Query Match:	31.44%	Indels:	5	
DB:	20	Gaps:	3	
US-09-513-999C-3792.COPY_51_161 (1-111) x AAY29918 (1-640)				
OY	110 GTACAGCACGACATCACTGCACAGACCAGCG-----AGTTTTCATCTCCGGG	57		
Db	292 IileInclnLurPro---GlnglurIrnrlnglnlgnlPrgluIlleusSerphethrlyrrmgily	310		
OY	56 CCCGAAGAAGTCCTGCCAATAAGATGAAAAAGATCCTGCCAAGCGCAAAGAT	9		
Db	311 Asprgrgser-----tlgsasplgluglucylsglcenlcnlgnlunsp	324		
RESULT 10				
ID	AAV29917 standard; Protein; 648 AA.			
AC	AAV29917;			
DE	18-NOV-1999 (first entry)			
DE	Human CHD1 protein encoded by transcript cDNA1.			
TW	Human; coronary heart disease susceptibility gene; CHD1; mutation;			

XX chromosome 11; diagnosis; screening; PCR primer; metabolic disorder;
KM detection; hypoadiphalipoproteinemia; familial combined hyperlipidaemia;
KM insulin resistant syndrome X; multiple metabolic disorder; obesity;
KM diabetes; dyslipidaemic hypertension.
XX
XX Homo sapiens.
XX OS
XX MO9945112-A2.
XX PN
XX 10-SEP-1999.
XX PD
XX PF 04-MAR-1999; 99WO-US04682.
XX PR 04-MAR-1998; 98US-0034941.
XX PR 06-APR-1998; 98US-0080934.
XX PA (MYRI-) MYRIAD GENETICS INC.
XX PI Ballinger DG, Ding W, Wagner S, Hess MA;
XX WP1: 1999-540844/45.
XX DR N-PSDB; AA226832.
XX
XX New isolated coronary heart disease susceptibility gene, used to
PT develop products for diagnosis and treatment of coronary heart disease
PT and metabolic disorders -
XX
XX Claim 11; Page 124; 297pp; English.
XX
XX The present invention describes the human chromosome 11-linked coronary
CC heart disease susceptibility gene (CHD1). Mutations in the CHD1 locus
CC in the genome are indicative of a predisposition to coronary heart
CC disease or to metabolic disorders related to lipid metabolism.
CC Products from the present invention can be used in the diagnosis
CC of predisposition to coronary heart disease and to metabolic disorders,
CC including hypoadiphalipoproteinemia, familial combined hyperlipidaemia,
CC insulin resistant syndrome X or multiple metabolic disorder, obesity,
CC diabetes and dyslipidaemic hypertension. CHD1 proteins can be used for
CC treating coronary heart disease and metabolic disorders. The products
CC can also be used for detection and drug screening. AA226832 to AA226841
CC and AA227027 to AA227029 represent human CHD1 nucleotide sequences.
CC AA229917 to AA229926 represent human CHD1 proteins and protein sequences
CC used in the exemplification of the present invention. AA226842 to
CC AA226862 represent primers used in the identification of human CHD1;
CC AA226863 to AA227014 represent PCR primers used in the screening of
CC mutations in human CHD1; AA227015 to AA227026 represent oligonucleotides
CC used in the exemplification of the present invention.
XX
XX Sequence 648 AA;
SQ
Alignment Scores:
Pred. No.: 9.33 Length: 648
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5
DB: 20 Gaps: 3
US-09-513-999c-3792_COPY_51_161 (1-111) x AA229917 (1-648)
QY 110 GTAGAGCAGCCACTGACGACACAGACCAGG-----AGTTTACATCTCCGGC 57
DB 300 IIEGInGInPro---GInGInuThrGInGInuProGInuIleuSerPheThrTYrThrly 318
QY 56 CCCAGAGTCCCTGCAGAAAGATGAAAAAGAAATCTCGAAGCAGCAAAAGAT 9
DB 319 AsPArGser-----LysAspGInGInGInuGInuGInuGInuAsp 332
RESULT 11
AAB95518
ID AAB95518 standard; Protein; 648 AA.
AC AAB95518;

XX
XX 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:18091.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS
XX Homo sapiens.
XX PN
XX EP1074617-A2.
XX PD
XX 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELT-) HELIX RES INST.
XX PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WP1: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 18091; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 648 AA;
SQ
Alignment Scores:
Pred. No.: 9.33 Length: 648
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5
DB: 22 Gaps: 3
US-09-513-999c-3792_COPY_51_161 (1-111) x AAB95518 (1-648)
QY 110 GTAGAGCAGCCACTGACGACACAGACCAGG-----AGTTTACATCTCCGGC 57
DB 300 IIEGInGInPro---GInGInuThrGInGInuProGInuIleuSerPheThrTYrThrly 318

QY 56 CCAGAGTCCCTGCAAGATGAAAGAAATCTGCAAGCAAAAGAT 9
||||| : : : : :
Db 319 AsparGSer-----LysAspGluGluGluCysLeuGluGlnIuasp 332

RESULT 12
ABG22576
ID ABG22576 standard; Protein; 1176 AA.
XX
AC ABG22576;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22567.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI: 2001-639362/73.
DR N-PSDB; AAS86763.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20: SEQ ID NO 52935; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1176 AA;

Alignment Scores:
Pred. No.: 9.76 Length: 1176
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5

DB: 22 Gaps: 3
US-09-513-999c-3792_COPY_51_161 (1-111) x ABG22576 (1-1176)

QY 110 GTAGAGACGCCCTGACGACACACAGAGACCCAGG-----AGTTTACATACTCCGCG 57
: : : : :
Db 352 lIeGInGluPro--GlnGluThrGInGluProGluIleLeuSerPheThrTyThrGly 370

QY 56 CCAGAGTCCCTGCAAGATGAAAGAAATCTGCAAGCAAAAGAT 9
||||| : : : : :
Db 371 AsparGSer-----LysAspGluGluGluCysLeuGluGlnIuasp 384

RESULT 13
AAM24448
ID AAM24448 standard; Protein; 52 AA.
XX
AC AAM24448;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1973.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI: 2001-476164/51.
DR N-PSDB; AAM99107.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 20: Page 1256; 1275bp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 52 AA;

Alignment Scores:
Pred. No.: 16.3 Length: 52
Score: 61.00 Matches: 13
Percent Similarity: 45.95% Conservative: 4
Best Local Similarity: 35.14% Mismatches: 14
Query Match: 29.33% Indels: 6
DB: 22 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x AAM24448 (1-52)

QY 9 ATCTTTTCCTTCGAGGACTC-----TTTTCATCTTTGAGGAGACT 50
 ::||| ||| ||||| ||| |||:::
 Db 6 ValhelaValIleuCySGIleuPheLeuTyLeuCySPhenIleuPhePheSerVal 25
 QY 51 TCTGGGGCCGAGTATGTAACACTCTGCTCTGTGTGCTGCTGAGTGG 101
 ::||| ||||| ::||| ||| ||| |||
 Db 26 ThrGlnaIaGlyValSerGluProArgSerSerIaGlyThrProIaArg 42
 RESULT 14
 AAB48057
 ID AAB48057 standard; protein: 77 AA.
 XX
 AC AAB48057;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE Human extracellular signaling molecule (EXCS) (ID 1288847CD1).
 XX
 KW Extracellular signaling molecule; EXCS; anti-inflammatory; human;
 KW immunosuppressive; cytosolic; neuroprotective; gastrointestinal;
 KW vitruicide; antibacterial; anti-HIV; human immunodeficiency virus;
 KW antiinfectility; cerebroprotective; nootropic; antitumor; antifungal;
 KW anticonvulsant; tranquilizer; neuroleptic; vasotropic; gynecological;
 KW keratolytic; protozoacide; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN NC0200070049-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 19-MAY-2000; 2000MO-US13975.
 XX
 PR 19-MAY-1999; 99US-0134949.
 PR 15-JUL-1999; 99US-0144270.
 PR 30-JUL-1999; 99US-0146700.
 PR 04-OCT-1999; 99US-0157508.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Tang YT, Yue H, Lal P, Burford N, Bandman O, Baughn MR;
 PI Azimzal Y, Lu DAM, Patterson C;
 PI
 DR WPI: 2001-025021/03.
 DR N-PSDB; AAC84293.
 XX
 PT New human extracellular signaling nucleic acids and polypeptides useful
 PT for diagnosing, treating and preventing infections and
 PT gastrointestinal, neurological, reproductive, and
 PT autoimmune/inflammatory disorders -
 XX
 PS Claim 1; Page 85; 114pp; English.
 XX
 CC The invention provides human extracellular signaling molecules (EXCS)
 CC and polynucleotides which identify and encode EXCS. EXCS can be
 CC expressed by standard recombinant methodology. The amino acid and nucleic
 CC acid sequences of EXCS are useful for diagnosing, treating and
 CC preventing infections and gastrointestinal (peptic ulcer, dysphagia,
 CC pancreatitis), neurological (e.g. epilepsy, ischemic cerebrovascular
 CC disease, stroke), reproductive (infertility, ovulatory defects,
 CC endometriosis), autoimmune (inflammatory (actinic keratosis, acquired
 CC immunodeficiency syndrome (AIDS), Addison's disease), and cell
 CC proliferative disorders including cancers (of the breast, adrenal gland,
 CC bone). They may also be used to treat fatal familial insomnia,
 CC nutritional and metabolic diseases of the nervous system, myopathies,
 CC mental disorders (anxiety, schizophrenia, mood), as well as infections
 CC caused by parasites (malaria, leishmania, trypanosoma), viral
 CC (adenovirus, coronavirus, flavivirus), bacterial (e.g. pneumococcus,
 CC staphylococcus, bacillus), and fungal (aspergillus, blastomycosis,
 CC dermatophytes) agents. The nucleic acids, polypeptides, antagonists,
 CC agonists, pharmaceutical compositions, and antibodies may also be used
 CC for treating or preventing disorders associated with increased or

CC decreased expression or activity of EXCS. EXCS polynucleotides may also
 CC be used to detect and quantify gene expression in biopsied tissues in
 CC which expression of EXCS may be correlated with the disease, to determine
 CC presence or excess expression of EXCS, to monitor regulation of EXCS
 CC levels during therapeutic intervention, to detect the presence of
 CC associated disorders, as targets in microarray, to generate hybridization
 CC probes, and to detect differences in gene sequences among normal, carrier
 CC or affected individuals. Antibodies may also be used in diagnosing
 CC disorders, in monitoring patients being treated with EXCS agonists,
 CC antagonists or inhibitors. Sequences AAB48057-B48082 represent the EXCS
 CC of the invention.
 XX
 SQ Sequence 77 AA:
 XX
 Alignment Scores:
 Pred. No.: 19.4 Length: 77
 Score: 60.50 Matches: 13
 Percent Similarity: 59.26% Conservative: 3
 Best Local Similarity: 48.15% Mismatches: 10
 Query Match: 29.09% Indels: 1
 DB: 22 Gaps: 1
 US-09-513-999c-3792_copy_51_161 (1-111) x AAB48057 (1-77)
 QY 28 TCTTTTCATCTTCGAGGAGACTCTG---GGGCGGAGTATGTAACACTCTGCTGCTC 84
 ::||| ||| ||||| ||| ||| |||:::
 Db 23 ThrPheLeuGlySerGlnIleuLeuGlnSerProSerTyGluIuValGlyCys 42
 QY 85 TGTGTGTGCTGCTGAGTGGCTGC 105
 ||||| |||||
 Db 43 CysLeuMetThrArgGlyCys 49
 RESULT 15
 AAR79164
 ID AAR79164 standard; Protein: 218 AA.
 XX
 AC AAR79164;
 XX
 DT 22-DEC-1995 (first entry)
 XX
 DE Partial sequence of human alpha 3 chain of type IV collagen.
 XX
 KW Type IV collagen; alpha 3 chain; Alport syndrome; COL4A3 gene.
 XX
 OS Homo sapiens.
 XX
 PN US5424408-A.
 XX
 PD 13-JUN-1995.
 XX
 PF 30-NOV-1990; 90US-0621091.
 XX
 PR 30-NOV-1990; 90US-0621091.
 XX
 PA (UNITV) UNITV KANSAS MEDICAL CENT.
 PA (UYVA) UNITV YALE.
 XX
 PI Hudson BG, Morrison KE, Reeders ST;
 PI
 DR WPI: 1995-262631/34.
 DR N-PSDB; AAQ96291.
 XX
 PT cDNAs encoding human or bovine alpha-3 type 4 collagen peptide(s) -
 PT useful for detection and therapeutic removal of antibodies associated
 PT with Goodpasture syndrome
 XX
 PS Disclosure: Columns 7-10; 33pp; English.
 XX
 CC Using the PCR with primers derived from each end of the known 27 AA
 CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp
 CC bovine genomic fragment was amplified. This fragment was then used
 CC to a bovine lens cDNA library and a 1.5 kb partial cDNA clone was
 CC obtd. (clone KMC15). This encodes 238 residues of the triple helical

CC collagenous domain and all 233 residues of the C-terminal non-
CC collagenous (NC1) domain of the alpha 3 (IV) chain. This bovine
CC cDNA clone was used to screen a human kidney cDNA library and a
CC 2.7 kb human cDNA clone (clone KMC27) was obtd. This clone encodes
CC 218 residues of the NC1 domain and a portion of the 3' UTR region
CC of the human alpha 3(IV) chain. The COL4A3 gene localises to
CC chromosome 2 and therefore mutations in COL4A3 cannot be
CC responsible for Alport syndrome which is X-linked. An isolated
CC and substantially pure nt. having the sequence in AAQ96291
CC is claimed.
CC
XX

SQ Sequence 218 AA;

Alignment Scores:

Pred. No.:	21	Length:	218
Score:	60.50	Matches:	17
Percent Similarity:	56.76%	Conservative:	4
Best Local Similarity:	45.95%	Mismatches:	11
Query Match:	29.09%	Indels:	5
B:	16	Gaps:	3

US-09-513-999c-3792_COPY_51_161 (1-111) x AAR79164 (1-218)

```
QY 3 GGTGGATCTTTGCTTGCAGATCTTTTCATCTT-----GCAGGGACTCTT 53
    |||||
Db 121 GlyTrpIleSerLeuTrpLysGlyPheSerPheIleMetPheThrSerAlaGlySerGlu 140
    |||||
QY 54 GGGCCCGGA---GTATGTAAACTCCTGGCTCTGTGTGTGCGCTGAGTGG 101
    |||||
Db 141 GlyAlaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGlnGluPhe 156
    |||||
```

Search completed: April 30, 2003, 14:38:40
Job time : 34.5 secs